

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Міністерство освіти і науки України
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

Котвицький Ігор Валерійович

УДК 621.391.83

ДИСЕРТАЦІЯ

Оцінювання якості спотворених мовних та музичних сигналів

05.09.08 – Прикладна акустика та звукотехніка

Електрична інженерія

Подається на здобуття наукового ступеня кандидата технічних наук.

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

(підпис, ініціали та прізвище здобувача)

Науковий керівник Продеус Аркадій Миколайович, д.т.н., проф.

Київ – 2020

АНОТАЦІЯ

Котвицький І. В. Оцінювання якості спотворених мовних та музичних сигналів. - На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.09.08 – прикладна акустика та звукотехніка. – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», МОН України, Київ, 2020.

Дисертаційна робота присвячена об'єктивізації та автоматизації оцінювання якості мовних та музичних сигналів, спотворених в комунікаційних каналах. Високі вимоги до людських, часових та фінансових ресурсів є суттєвими недоліками суб'єктивного оцінювання. Використання об'єктивних (інструментальних) методів та систем оцінювання має дозволити позбутися вказаних недоліків. Крім того, актуальною проблемою є автоматизація суб'єктивного оцінювання якості мовних та музичних сигналів, оскільки від наявності та достовірності еталонних результатів суб'єктивного оцінювання залежить принципова можливість практичного використання об'єктивних мір якості та розбірливості мови. Висока складність та вартість наявних підходів до автоматизації суб'єктивного оцінювання робить актуальним пошук шляхів до створення простих, недорогих та надійних систем автоматизації суб'єктивного оцінювання якості мовних та музичних сигналів. Перехід від одноканальної до багатоканальної (просторово-часової) обробки акустичних сигналів породжує проблему порівняння різних алгоритмів обробки сигналів та вибору найкращого за критерієм якості результуючого сигналу, алгоритму. Таким чином, розробка нових та вдосконалення наявних методів оцінювання якості мовних та музичних сигналів є актуальною науково-технічною задачею акустичної експертизи комунікаційних каналів, що має важливе прикладне значення.

Автором проведено розробку нових та вдосконалення відомих методів оцінювання якості мовних та музичних сигналів, що забезпечують підвищення

достовірності та зменшення ресурсоемності акустичної експертизи трактів передачі звукової інформації.

Особливу увагу приділено аналізу та порівнянню поведінки таких об'єктивних показників якості мовних та музичних сигналів, як сегментне відношення сигнал-шум SSNR, лог-спектральні спотворення LSD, барк-спектральні спотворення BSD, перцептуальна оцінка якості мовного сигналу PESQ та перцептуальна оцінка якості звукового сигналу PEAQ, які є перспективними для використання в системах корекції сигналів в реальному масштабі часу. Ці дослідження є актуальними з точки зору стандартизації вимог до якості сучасних каналів передачі мовних та музичних сигналів в системах мобільного зв'язку, мережевих комунікацій, радіомовлення та телебачення, а також в приміщеннях різного призначення.

Виконано уточнюючий аналіз особливостей об'єктивних мір SSNR і LSD якості мовних і музичних сигналів на прикладі сигналів, обмежених за смугою частот, що дозволило дати пояснення виявленим особливостям вказаних мір і виробити ряд рекомендацій, спрямованих на їх практичне використання. Уточнено висновки щодо підвищеної чутливості сегментного відношення сигнал-шум SSNR до похибки вирівнювання сигналів та запропоновано рекомендації зі зниження такої чутливості. Зокрема, показано, що інтерполяція сигналів, що тестуються, дозволяє підвищити достовірність оцінювання з використанням показника SSNR, а саме: для мовних сигналів частоту дискретизації інтерпольованих сигналів достатньо збільшити вдвічі, тоді як для музичних сигналів необхідне збільшення у 5 разів. Крім того, показано, що можливими причинами порушення монотонності залежності LSD від ширини смуги частот є нормування сигналів, що порівнюються, за дисперсією, висока чутливість LSD до особливостей спектру сигналу, а також відмінність форми АЧХ використаних НЧ фільтрів від прямокутної форми.

Автором вперше отримано для реальних мовних та музичних сигналів кількісні суб'єктивні та об'єктивні оцінки ступеню впливу нелінійності фазової

характеристики тракту на якість мовних та музичних сигналів та встановлено, що слухова система людини є більш чутливою до фазових спотворень мовних сигналів, порівняно з музичними сигналами, а також є чутливою до знаку різниці часу групової затримки на низьких та високих частотах.

Вперше побудовано карти відповідності між суб'єктивною мірою якості DMOS та об'єктивними мірами SSNR, LSD, BSD та PEAQ для ситуацій лінійного спотворення сигналів (обмеженість смуги частот та нелінійність фазової характеристики тракту) та нелінійного їх спотворення (кліпування), що робить можливим калібрування відповідних програмно-апаратних систем об'єктивного оцінювання якості музичних сигналів. Таке калібрування, на відміну від використання коефіцієнта кореляції, є більш точним, оскільки припускає існування нелінійної залежності між суб'єктивними та об'єктивними оцінками якості сигналів.

Отримані автором залежності суб'єктивних оцінок якості кліпованих мовних та музичних сигналів від ступеня їх кліпування створюють необхідну базу для подальшого калібрування об'єктивних мір ступеня кліпування. Вперше запропоновано об'єктивні міри ступеня кліпування мовних та музичних сигналів у вигляді ексцесу, оберненої до нього величини, а також кореня квадратного із величини, оберненої до ексцесу.

Уточнено межі застосування об'єктивних оцінок якості мови замість суб'єктивних оцінок розбірливості зашумленої мови, що дозволяє (за певних значень відношення сигнал-шум понад 0 дБ) значно спростити процедуру об'єктивного оцінювання розбірливості мови.

Вдосконалено технологію автоматизації суб'єктивного оцінювання розбірливості мови артикуляційним методом, дієвість якої перевірено за допомогою розробленого прототипу автоматизованої системи артикуляційних випробувань каналів передачі української мови. Надано рекомендації щодо подальшого покращення розробленої системи, яку можна використовувати в якості прототипу при створенні національних стандартів оцінювання розбірливості мови в каналах зв'язку та в приміщеннях.

Уточнено результати зіставлення алгоритмів обробки музичних сигналів в багатомікрофонних масивах шляхом об'єктивного та суб'єктивного оцінювання якості звуку. Порівняння двох алгоритмів просторово-часової обробки за критерієм якості музичного сигналу, спотвореного адитивними завадами, дозволяє рекомендувати простіший, з точки зору обчислень, алгоритм суми та затримок замість більш складного конкурентного алгоритму.

Практичне значення отриманих результатів полягає у встановленні умов використання порівняно простих, з обчислювальної точки зору, об'єктивних мір якості сигналів замість складних перцептуальних мір; у встановленні принципової можливості використання коефіцієнту ексцесу та його простих функціональних перетворень в ролі об'єктивних мір якості сигналів, спотворених кліпуванням, що дозволяє спростити та підвищити надійність процедури виявлення кліпування; визначенні границь можливого застосування порівняно простих, з обчислювальної точки зору, методів оцінювання якості мовних сигналів замість більш складних методів оцінювання розбірливості мови; розробці комплексу комп'ютерних програм для автоматизації артикуляційних випробувань каналів передачі української мови в умовах дії шумової завади та реверберації. Отримані автором результати можуть використовуватися при акустичній експертизі приміщень та ліній зв'язку, а також в навчальному процесі вищих навчальних закладів України для підготовки інженерів-акустиків. Результати дисертаційної роботи впроваджені у практику оцінювання якості звукоізоляції шляхом оцінювання розбірливості мови у вигляді математичного забезпечення систем оцінювання розбірливості мови для науково-виробничого підприємства «ТОВ Глобал Акустікс», а також в учбовий процес у курсах лекцій кафедри акустики та акустoeлектроніки з дисципліни «Комп'ютерна обробка акустичних сигналів» у розділах «Оцінювання якості акустичних сигналів», «Автоматизація оцінювання розбірливості мови», а також з дисципліни «Захист акустичної інформації» у розділі «Прогнозування та вимірювання розбірливості мови» (№ДР 0116U008959).

Впровадження результатів дисертаційної роботи підтверджено відповідними актами.

Ключові слова: фазові спотворення, кліпований мовний сигнал, кліпований музичний сигнал, суб'єктивна оцінка, якість мовного сигналу, якість музичного сигналу, показники якості, карта відповідності, розбірливість мови, артикуляційні випробування, шум, реверберація, односкладові звукосполучення, масиви мікрофонів.

SUMMARY

Kotvytskyi I. V. Assessing the quality of distorted speech and music signals. - Manuscript.

Thesis for a Candidate degree in Technical Science, Specialty 05.09.08 Applied Acoustics and Audio Engineering. - National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ministry of Education and Science of Ukraine, Kyiv, 2020.

The thesis is devoted to the objectification and automation of evaluation of the quality of speech and music signals distorted in communication channels. High requirements for human, time and financial resources are the significant drawbacks of subjective evaluation. The use of objective (instrumental) assessment methods and systems should help to eliminate these shortcomings. In addition, the actual problem is the automation of subjective assessment of the quality of speech and music signals, since the availability and reliability of reference results of subjective evaluation depends on the fundamental possibility of practical use of objective measures of quality and intelligibility of speech. The high complexity and cost of existing approaches to subjective automation makes it important to find ways to create simple, inexpensive and reliable systems for subjective evaluation of the quality of speech and music signals. The transition from single-channel to multi-channel (spatio-temporal) processing of acoustic signals raises the problem of comparing different algorithms of signal processing and choosing the best algorithm for quality of the resulting signal. Thus, the development of new and the refinement of existing methods for evaluating

the quality of speech and music signals is an urgent scientific and technical task of acoustic examination of communication channels, which has significant practical value.

The author has developed new and improved some of the known methods of evaluating the quality of speech and musical signals, which provide increased accuracy and reduced the intensity of acoustic expertise of the paths of transmission of audio information.

Particular attention is paid to the analysis and comparison of the behavior of such objective indicators of the quality of speech and music signals, such as segmental signal-to-noise ratio SSNR, log-spectral distortion of LSD, bar-spectral distortion of BSD, perceptual assessment of the quality of the speech signal PESQ and perceptual evaluation of the quality of the audio signal PEAQs that are promising for use in real-time signal correction systems. These studies are relevant in terms of standardizing the requirements for the quality of modern voice and music transmission channels in mobile communications, network communications, radio and television, as well as in various applications.

A refinement of the features of the SSNR and LSD objective measures of the quality of the speech and music signals on the example of signals limited by the frequency band was performed, which allowed to give an explanation of the revealed features of these measures and to make a number of recommendations aimed at their practical use. Conclusions on the increased sensitivity of the segmental signal-to-noise ratio SSNR to the error of signal alignment have been clarified and recommendations for reducing such sensitivity have been proposed. In particular, it has been shown that the interpolation of the signals being tested improves the accuracy of the estimation using the SSNR, namely that for the speech signals, it is sufficient to double the sampling rate of the interpolated signals, whereas for the musical signals the sampling rate should be increased by 5 times. It is shown that the possible reasons for the disruption of the monotonicity of the LSD dependence on the bandwidth are the normalization of the signals being compared, the variance, the high sensitivity of the

LSD to the features of the signal spectrum, and the difference in the frequency response Rectangular LF filters of rectangular shape.

The author first obtained quantitative subjective and objective estimates of the degree of nonlinearity of the phase characteristics of the tract on the quality of speech and music signals for real speech and music signals, and found that the human auditory system is more sensitive to phase distortions of speech signals compared to musical signals , and is also sensitive to the difference between low and high bandwidth delays.

For the first time maps of correspondence between subjective DMOS quality measure and objective SSNR, LSD, BSD and PEAQ measures for situations of linear signal distortion (frequency band limitation and nonlinearity of phase characteristic of the path) and nonlinear distortion (clipping) are constructed, which makes calibration possible appropriate software and hardware systems for objective evaluation of the quality of music signals. This calibration, in contrast to the use of the correlation coefficient, is more accurate because it assumes a nonlinear relationship between subjective and objective estimates of signal quality.

Dependences of subjective quality scores on clipped speech and music cues on the degree of their clipping provides the necessary basis for further calibration of the objective measures of the degree of clipping obtained by the author. For the first time, objective measures of the degree of clipping of linguistic and musical signals into the kinds of excess, inverted to its magnitude, and the square root of magnitude, inverted to excess, are proposed.

The limits of the use of objective speech quality assessments instead of subjective noisy speech intelligibility have been clarified, which, for certain signal-to-noise ratios of more than 0 dB, can greatly simplify the procedure for objectively assessing speech intelligibility.

The technology of automation of subjective assessment of speech intelligibility by the articulation method has been improved, the efficiency of which was tested by means of the developed prototype of the automated system of articulation testing of the

Ukrainian speech transmission channels. Suggestions are made to further improve the developed system, which can be used as a prototype for the creation of national standards for the assessment of speech intelligibility in communication channels and in premises.

The results of comparison, by objective and subjective assessment of sound quality, of signals processing algorithms in the multi-microphone arrays are refined. Comparison of two algorithms of space-time processing based on the criterion of the quality of the music signal distorted by additive interference allows us to recommend a simpler, from the point of the calculations view, algorithm of the sum and delays instead of the more complicated competitive algorithm.

The practical significance of the results obtained lies in establishing conditions for the use of relatively simple objective measures of signal quality (in terms of computation) instead of complex perceptual measures; in establishing the fundamental possibility of using the coefficient of excess and its simple functional transformations as objective measures of the quality of signals distorted by clipping, which makes it possible to simplify and increase the reliability of the clipping detection procedure; defining the boundaries of the possible use of relatively simple methods for assessing the quality of speech signals (in terms of computation) instead of more sophisticated methods for assessing speech intelligibility; development of a set of computer programs for automation of Ukrainian speech transmission channels articulation testing in the conditions of noise interference and reverberation.

The results obtained by the author can be used in the acoustic examination of premises and communication lines, as well as in the educational process of higher educational establishments of Ukraine for the training of acoustic engineers.

The results of the thesis paper were implemented in the development of mathematical support for an automated system of subjective assessment of speech intelligibility, which was developed at the Department of Acoustics and Acoustoelectronics, NTUU "KPI named after Igor Sikorsky" at the Scientific and

Research Center "Objective and subjective assessment of speech quality and intelligibility (state registration No. 0116U008959), as well as introduced into the educational process in the courses of lectures at the Department of Acoustics and Acoustoelectronics in the course "Computer processing of acoustic signals", in the sections "quality of acoustic signals", "Automation of speech intelligibility assessment", as well as in the course "Protection of acoustic information", in the section "Prediction and measurement of speech intelligibility".

Keywords: phase distortions, clipped speech signal, clipped musical signal, subjective assessment, quality of speech signal, quality of musical signal, quality indicators, conformity map, speech intelligibility, articulation tests, noise, reverberation, single-component sound combinations, microphone arrays.

Список публікацій

1. И. В. Котвицкий, А. Н. Продеус, «Объективное и субъективное оценивание качества речевых и музыкальных сигналов, подвергнутых фазовым искажениям». - ISSN 1811-4512. *Электроника и связь*. 2016, № 2. С. 25-31.
Особистий внесок здобувача полягає в комп'ютерному моделюванні фазових спотворень та проведенні експерименту (Index Copernicus).
2. A.M.Prodeus, I.V. Kotvytskyi, A.A.Ditiashov, «Assessment of clipped speech quality», *Electronics and Control Systems*, 2018, N 4(58), P. 11-18. DOI:10.18372/1990-5548.58.13504. *Особистий внесок здобувача полягає в проведенні розрахунків, організації експерименту та аналізі результатів експерименту (Index Copernicus).*
3. А. Продеус, А. Вітик, О. Дворник, І. Котвицький, О. Чайка, М. Ярошенко, «Суб'єктивне оцінювання розбірливості мови на тлі шуму та реверберації». *Мікросистеми, електроніка та акустика*. Том 23, № 2, 2018. ISSN 2523-4447, DOI: 10.20535/2523-4455.2018.23.2.128820. С. 66-73.

Особистий внесок здобувача полягає в дослідженні комп'ютерної моделі системи артикуляційних досліджень та проведенні експерименту (Index Copernicus).

4. М.П. Юраков, І.В. Котвицький, А.М. Продеус, «Оцінювання якості музичних сигналів, обмежених за смугою частот». ISSN 2523-4447. Мікросистеми, Електроніка та Акустика, 2018, vol. 23, № 4,. DOI: 10.20535/2523-4455.2018.23.4.133785. С. 58-64. *Особистий внесок здобувача полягає в проведенні комп'ютерного моделювання, проведенні експерименту та аналізі результатів експерименту (Index Copernicus).*
5. A. Prodeus, K. Bukhta, P. Morozko, O. Serhiienko, I. Kotvytskyi, O. Dvornyk, «Automated Subjective Assessment of Speech Intelligibility in Various Listening Modes», ISSN 2523-4447. Microsystems, Electronics and Acoustics, 2018, vol. 23, no. 3, P.49-57. *Особистий внесок здобувача полягає в дослідженні комп'ютерної моделі системи артикуляційних досліджень та проведенні експерименту (Index Copernicus).*
6. I. Kotvytskyi, «Objective and Subjective Music Quality Assessment for Signals Limited by Frequency Band». IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF). м. Львів, 16-18 жовтня 2017 року.
7. А.Н. Продеус, В.С. Дидковский, И. В. Котвицкий, «Об оценивании качества речи с помощью показателя «логарифмические спектральные искажения»». Тези доповідей 5-ї міжнародної науково-технічної конференції «Інформаційні проблеми теорії акустичних, радіоелектронних і телекомунікаційних систем» IPST-2016. – Харків: НТУ «ХПІ», 2016. С. 14. *Особистий внесок здобувача полягає у розрахунку та аналізі об'єктивних мір якості сигналів.*
8. A. Prodeus, V. Didkovskyi, M. Didkovska, I. Kotvytskyi, «Automated Subjective Evaluation of Speech Intelligibility in Noise and Reverberation». *Proceedings of IEEE 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, February 20-24, 2018, Lviv-Slavske, Ukraine, P. 555-558. ISBN: 978-1-5386-

- 2557-6. DOI: 10.1109/TCSET.2018.8336263 *Особистий внесок здобувача полягає в дослідженні комп'ютерної моделі системи артикуляційних досліджень та проведенні експерименту.*
9. A. Prodeus, I. Kotvytskyi, «On Reliability of Log-Spectral Distortion Measure in Speech Quality Estimation», Proceedings of IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), P. 17-19. October 2017, Kyiv, Ukraine. *Внесок здобувача полягає у розрахунку та аналізі об'єктивних мір якості сигналів.*
10. A. Prodeus, V. Didkovskyi, M. Didkovska, I. Kotvytskyi, «On Peculiarities of Evaluating the Quality of Speech and Music Signals Subjected to Phase Distortion», *Proceedings of IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO)*, April 18-20, 2017, Kyiv, Ukraine. – P. 455-460. *Особистий внесок здобувача полягає в розробці та дослідженні комп'ютерних моделей.*
11. A. Prodeus, I. Kotvytskyi, A. Ditiashov, «Clipped Speech Signals Quality Estimation», Proceedings of 5th International Conference "Methods and Systems of Navigation and Motion Control" (MSNMC-2018), October 16-18, 2018, Kyiv, Ukraine, P. 151-155. ISBN: 978-1-5386-5869-7. DOI: 10.1109/MSNMC.2018.8576308. *Особистий внесок здобувача полягає в проведенні комп'ютерного моделювання та проведенні експерименту.*
12. A. Prodeus, K. Bukhta, P. Morozko, O. Serhiienko, I. Kotvytskyi, I. Shherbenko, «Automated System for Subjective Evaluation of the Ukrainian Speech Intelligibility», Proceedings of IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO), April 24-26, 2018 Kyiv, Ukraine, P. 533-538. ISBN: 978-1-5386-6382-0. *Особистий внесок здобувача полягає в дослідженні комп'ютерної моделі системи артикуляційних досліджень та проведенні експерименту.*
13. A. Prodeus, I. Kotvytskyi, M. Didkovska, «Kurtosis and Its Transformations as Objective Measures of Clipping Value and Speech Quality», Proceedings of IEEE 5th International Conference Actual Problems of Unmanned Aerial

Vehicles Developments (APUAVD), P. 22-24 October 2019, Kyiv, Ukraine.
Особистий внесок здобувача полягає в проведенні комп'ютерного моделювання.

- 14.A. Prodeus, M. Didkovska, I. Kotvytskyi, «Using Kurtosis for Objective Assessment of the Musical Signals Clipping Degree», Proc. 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), October 2019, Kyiv, Ukraine. *Особистий внесок здобувача полягає в проведенні комп'ютерного моделювання.*
- 15.A. Prodeus, I. Kotvytskyi, A. Garasiuk, A. Semenchuk, A. Telenga, S. Klymenko, V. Lozinskyi, O. Andriichenko, O. Denysenko, «Subjective Assessment of the Influence of Room Characteristics on Intelligibility of Noised Speech». Proceedings of IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO), P. 16-18 April 2019, Kyiv, Ukraine. DOI: 10.1109/ELNANO.2019.8783250. *Особистий внесок здобувача полягає в дослідженні комп'ютерної моделі системи артикуляційних досліджень та проведенні експерименту.*
- 16.Комп'ютерна програма «Автоматизована система суб'єктивного оцінювання розбірливості мови»: свідоцтво про реєстрацію авторського права на твір № 85027/І. В. Котвицький та ін. Україна: Міністерство економічного розвитку і торгівлі України; опубл. 30.01.19. *Особистий внесок здобувача полягає у тестуванні автоматизованої системи оцінювання розбірливості мови.*

ЗМІСТ

ВСТУП.....	17
1. Сучасний стан акустичної експертизи комунікаційних каналів.....	24
1.1. Спотворення мовних і музичних сигналів в комунікаційних каналах.....	24
1.2. Суб'єктивний і об'єктивний підходи до оцінювання якості мови і музики.....	27
1.2.1. Суб'єктивні міри якості мови і музики	27
1.2.2. Об'єктивні міри якості мови і музики	29
1.3. Суб'єктивний і об'єктивний підходи до оцінювання розбірливості мови	36
1.3.1. Суб'єктивний підхід до оцінювання розбірливості мови	37
1.3.2. Об'єктивний підхід до оцінювання розбірливості мови	38
1.4. Багатомікрофонні масиви в комунікаційних каналах	41
Висновки до розділу 1	46
2. Оцінювання якості мови і музики при лінійних і нелінійних спотвореннях.....	48
2.1. Суб'єктивне оцінювання якості мовних та музичних сигналів.....	48
2.2. Об'єктивне оцінювання якості мовних та музичних сигналів	49
2.2.1. Залежність результатів обчислень міри SSNR від похибки вирівнювання сигналів за часом.....	49
2.2.2. Особливості оцінки міри LSD	51
2.3. Об'єктивне оцінювання якості музичних сигналів, обмежених за смугою частот.....	54
2.3.1. Пілотні дослідження якості музичних сигналів, обмежених за смугою частот.....	55

2.3.2. Уточнюючі дослідження якості музичних сигналів, обмежених за смугою частот.....	60
2.4. Оцінка ступеня впливу фазових спотворень на якість мови і музики.	66
2.5. Оцінювання ступенів кліпування і якості кліпованої мови та музики. ..	76
2.5.1. Вступні зауваження	77
2.5.2. Параметризація оцінок якості сигналів при жорсткому двосторонньому кліпуванні	78
2.5.3. Експериментальні дослідження мір ступеня кліпування	81
2.5.4. Міри ступеня кліпування як міри якості кліпованих сигналів	85
Висновки до розділу 2	91
3. Оцінювання розбірливості мови в умовах дії шуму і реверберації	93
3.1. Автоматизація артикуляційних досліджень.....	93
3.2. Експериментальна перевірка дієздатності та ефективності автоматизованої системи артикуляційних досліджень.....	96
3.2.1. Перший етап експериментальних досліджень.....	97
3.2.2. Другий етап експериментальних досліджень	102
3.2.3. Суб'єктивна оцінка впливу характеристик приміщення на розбірливість мови спотвореної шумом	104
3.3. Оцінювання розбірливості мови через оцінювання якості мови.	108
3.3.1. Організація експериментальних досліджень	110
3.3.2. Результати суб'єктивного оцінювання розбірливості та якості мовних сигналів	111
3.3.3. Порівняння результатів об'єктивного оцінювання якості мовних сигналів із результатами суб'єктивного оцінювання їх якості та розбірливості	112
3.3.4. Висновки за отриманими результатами	114

Висновки до розділу 3	115
4. Зіставлення алгоритмів обробки музичних сигналів у багатомікрофонних масивах.....	117
4.1. Комп'ютерна модель багатомікрофонного масиву.....	117
4.2. Суб'єктивне оцінювання якості музики із застосуванням мікрофонних масивів.....	119
4.2.1. Пілотні експериментальні дослідження з оцінювання якості музичних сигналів у багатомікрофонних системах.....	121
4.2.2. Уточнюючі експериментальні дослідження з оцінювання якості музичних сигналів в багатомікрофонних системах	130
Висновки до розділу 4	139
ВИСНОВКИ	140
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	143
Додаток 1. Довідки про впровадження результатів роботи	152
Додаток 2. Таблиця артикуляційних звукосполучень	154
Додаток 3. Лістинги програм автоматизованої системи суб'єктивного та об'єктивного оцінювання якості та розбірливості мовних та якості музичних сигналів в Matlab	155
Додаток 4. Свідоцтво про авторське право на комп'ютерну програму .	161

ВСТУП

Обґрунтування теми дослідження. Акустична експертиза каналів передачі інформації, а також точність, простота, швидкість і вартість цієї експертизи є надзвичайно важливими з огляду на роль технічних засобів комунікації в сучасному суспільстві.

Оцінювання якості мовних та музичних сигналів, спотворених в каналах передачі інформації, є основною складовою акустичної експертизи. Перевагою суб'єктивного оцінювання якості сигналів є еталонний характер отриманих результатів, оскільки якість звукових сигналів оцінюється кінцевими користувачами. Суттєвим недоліком суб'єктивного оцінювання є високі вимоги до фінансових, часових та людських ресурсів. Такого недоліку позбавлені об'єктивні (інструментальні) методи й системи оцінювання, чим і пояснюється незмінний інтерес науковців та інженерів до таких систем. Значний внесок в розвиток об'єктивних методів оцінювання якості мовних сигналів зробили такі вчені як Покровський Н.Б., Биков Ю.С., Сапожков М.А., Калінцев Ю.К., Фант Г., Рабінер Л., Янг В., Цвіккер Е., Фельдкеллер Р., Луізоу П., Мюллер С., Коте Т., Квакенбуш І., Бірендс Й., Блаурт Й., Бредлі Д., Арвейлер І. та інші.

Суттєвою проблемою при застосуванні методів об'єктивного оцінювання є різноманітність причин спотворення акустичних сигналів. Це дія шумових та ревербераційних завад, акустичне та електронне відлуння, нерівномірність амплітудної частотної характеристики та обмеженість смуги частот тракту, нелінійність фазочастотної характеристики тракту, кліпування сигналів, спотворення сигналів при використанні складних алгоритмів їх обробки, спрямованих на компресію сигналів, придушення шуму, реверберації, відлуння, корекцію передатних характеристик тракту, просторову вибіркковість тощо.

Іншою проблемою є надзвичайно велика кількість – більше двох тисяч – об'єктивних мір якості сигналів, що значно ускладнює вибір потрібної міри. Однією з причин такого різноманіття є те, що результати використання простих для обчислень об'єктивних мір не завжди добре узгоджуються із результатами суб'єктивного оцінювання. Це стимулює дослідників ускладнювати об'єктивні міри, щоб досягти кращого врахування особливостей слухової системи людини. При цьому, на жаль, залишаються не до кінця дослідженими потенційні можливості простих для обчислень об'єктивних мір, зокрема, для сегментного відношення сигнал-шум та лог-спектральних спотворень.

Нагальною проблемою є отримання еталонних результатів суб'єктивного оцінювання якості та розбірливості мови – від наявності та достовірності таких еталонів залежить принципова можливість практичного використання об'єктивних мір якості та розбірливості мови. Проте отримання достовірних еталонних результатів суб'єктивного оцінювання потребує обробки великих обсягів звукового матеріалу, що призводить до необхідності автоматизувати процедуру артикуляційних випробувань. На жаль, системи автоматизації артикуляційних випробувань, що використовувалися такими вченими як Бредлі Д. та Арвейлер І., є надто складними та дорогими. Тому актуальним є питання створення простих, недорогих та надійних систем автоматизації оцінювання якості та розбірливості мови.

Важливою особливістю мовних сигналів є те, що якість та розбірливість мови не є тотожними поняттями: якісний сигнал не завжди є розбірливим, а розбірливий сигнал не завжди є якісним. Оскільки оцінювання розбірливості мови об'єктивними методами є набагато складнішим, порівняно з оцінюванням якості мовних сигналів, виникає питання про пошук умов, за яких можливе використання простих мір якості замість складних мір розбірливості мови. На жаль, це важливе питання й досі залишається недостатньо дослідженим.

Недостатньо вивченим на теперішній час є також питання суб'єктивного та об'єктивного оцінювання якості музичних сигналів. Праці таких дослідників як

Алдошина І., Прітте Р., Гянакопулос Т., Пікракіс А. переважно стосуються оцінювання параметрів музичних сигналів, а питання об'єктивного оцінювання якості спотворених музичних сигналів в них майже не висвітлено. Цей недолік в певній мірі компенсовано в працях Бірендс Д., Тіеде Т., Хюбер Р., Барбедо Д. та інших, де запропоновано так звані «перцепційні міри» якості акустичних сигналів, що містять різні моделі слухової системи людини. Складність таких мір та високі вимоги до обчислювальних ресурсів є їхнім суттєвим недоліком, що мотивує до пошуків більш простих і водночас надійних технічних рішень.

Оскільки оцінювання якості мовних та музичних сигналів є основою сучасних цифрових систем акустичної експертизи комунікаційних каналів, створення відповідного математичного та програмного забезпечення є **актуальною** науково-технічною задачею, що має важливе прикладне значення. Цим визначається актуальність та практичне значення теми дисертаційного дослідження.

Зв'язок роботи з науковими планами, програмами, темами. Дисертаційну роботу виконано на кафедрі акустики та акустoeлектроніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» в рамках НДР «Об'єктивне та суб'єктивне оцінювання якості та розбірливості мовлення» (№ДР 0116U008959).

Мета і завдання дослідження. Метою роботи є дослідження методів оцінювання якості мовних і музичних сигналів, підданих лінійним та нелінійним спотворенням, для подальшого вдосконалення даних методів.

Для досягнення цієї мети необхідно знайти розв'язок наступних завдань:

1. Суб'єктивне оцінювання якості мовних і музичних сигналів, спотворених в трактах передачі й перетворення інформації, що дозволяє отримати еталонні характеристики слухової системи людини в умовах сприйняття спотворених сигналів.

2. Автоматизація суб'єктивного оцінювання розбірливості мови, що дозволяє істотно зменшити трудовитрати при збереженні високої достовірності результатів артикуляційних випробувань.
3. Зіставлення потенційних можливостей об'єктивних мір якості мовних та музичних сигналів, що дозволяє здійснювати обґрунтований вибір кращої міри за такими показниками як достовірність та швидкість акустичної експертизи.

Об'єктом дослідження є процеси обробки мовленнєвих та музичних сигналів, спрямовані на визначення якості сигналів, спотворених в каналах передачі акустичної інформації.

Предметом дослідження є методи обробки спотворених сигналів в системах суб'єктивної та об'єктивної акустичної експертизи.

Методи дослідження. При виконанні досліджень використані методи теорії цифрової обробки сигналів, методи оцінювання розбірливості мови, методи оцінювання якості акустичних сигналів, методи просторово-часової обробки акустичних сигналів. При цьому, для опису процесів оцінювання якості мовних та музичних сигналів, що відбуваються при сприйнятті акустичних сигналів слуховою системою людини, використано аналітичні та комп'ютерні моделі слухової системи. Достовірність отриманих результатів підтверджується зіставленням одержаних результатів з опублікованими в науковій літературі відповідними результатами, в тому числі експериментальними.

Наукова новизна отриманих результатів.

1. Вперше для реальних мовних та музичних сигналів встановлено, що слухова система людини є чутливою до знаку різниці часу групової затримки для низьких та високих частот, а також отримано суб'єктивні та об'єктивні оцінки ступеню впливу нелінійності фазової характеристики тракту на якість сигналів.

2. Вперше запропоновано використання ексцесу та його функціональних перетворень як об'єктивних мір якості мовних та музичних сигналів, спотворених кліпуванням.
3. Вперше вказано на підвищену чутливість до спектральних властивостей сигналів простої в обчислювальному плані об'єктивної міри якості сигналів у вигляді лог-спектральних спотворень.
4. Знайшло подальший розвиток положення щодо підвищеної чутливості сегментного відношення сигнал-шум до похибки вирівнювання сигналів та запропоновано рекомендації зі зниження такої чутливості.
5. Знайшло подальший розвиток зіставлення, шляхом об'єктивного та суб'єктивного оцінювання якості звуку, алгоритмів обробки музичних сигналів в багатомікрофонних масивах.

Практичне значення одержаних результатів полягає в наступному:

1. Встановлено принципову можливість використання, при акустичній експертизі комунікаційних каналів, порівняно простих, з обчислювальної точки зору, об'єктивних мір якості сигналів SSNR, LSD та BSD замість складних перцепційних мір PESQ та PEAQ.
2. Встановлено принципову можливість використання коефіцієнту ексцесу та його простих функціональних перетворень як об'єктивних мір якості мовних та музичних сигналів, спотворених кліпуванням, що дозволяє спростити та підвищити надійність процедури виявлення кліпування.
3. Визначено межі можливого застосування порівняно простих, з обчислювальної точки зору, методів оцінювання якості мовних сигналів замість більш складних методів оцінювання розбірливості мови.
4. Розроблено комплекс комп'ютерних програм для автоматизації артикуляційних випробувань каналів передачі української мови в умовах дії шумової завади та реверберації, що дозволяє значно знизити вимоги до фінансових, часових та людських ресурсів при проведенні артикуляційних випробувань.

5. Виконано порівняння двох алгоритмів просторово-часової обробки, таких як DS та LCMV, за критерієм якості музичного сигналу, спотвореного адитивними завадами. Показано, що простіший для обчислення алгоритм DS дозволяє отримати якісніші сигнали.

Результати дисертаційної роботи впроваджені у практику оцінювання якості звукоізоляції шляхом оцінювання розбірливості мови у вигляді математичного забезпечення систем оцінювання розбірливості мови для науково-виробничого підприємства «ТОВ Глобал Акустікс», а також в учбовий процес у курсах лекцій кафедри акустики та акустoeлектроніки з дисципліни «Комп'ютерна обробка акустичних сигналів» у розділах «Оцінювання якості акустичних сигналів», «Автоматизація оцінювання розбірливості мови» та дисципліни «Захист акустичної інформації» у розділі «Прогнозування та вимірювання розбірливості мови» (№ДР 0116U008959). Впровадження результатів дисертаційної роботи підтверджено відповідними актами.

Особистий внесок здобувача відображений публікаціями [1-16]. У наукових працях, виконаних у співавторстві, здобувачу належить: в роботах [1, 10] комп'ютерне моделювання фазових спотворень, проведення експерименту та аналіз отриманих результатів, в роботах [2, 4, 6, 7, 9, 11, 13, 14] – організація експерименту та аналіз результатів експерименту, в роботах [3, 5, 8, 12, 15, 16] – дослідження комп'ютерної моделі системи артикуляційних досліджень та проведенні експерименту.

Робота є результатом самостійних досліджень Котвицького І. В.

Апробація результатів дисертації. Основні положення та результати дисертації обговорювались на:

- V Міжнародній науково-технічній конференції молодих вчених «Інформаційні проблеми теорії акустичних, радіoeлектронних і телекомунікаційних систем» (IPST-2016), 29.11-1.12, 2016, Харків;

- Міжнародній науково-технічній конференції молодих вчених «International Young Scientists Forum on Applied Physics and Engineering (YSF)», Львів, 17.10.2017;

- V International Conference "Methods and Systems of Navigation and Motion Control" (MSNMC-2018), October 16-18, 2018, Kyiv, Ukraine;

- V International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), October 17-19, 2017, Kyiv, Ukraine;

- 37th International Conference on Electronics and Nanotechnology (ELNANO), April 18-20, 2017, Kyiv, Ukraine;

- 38th International Conference on Electronics and Nanotechnology (ELNANO), April 24-26, 2018 Kyiv, Ukraine;

- 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), February 20-24, 2018, Lviv-Slavske, Ukraine;

- 39th International Conference on Electronics and Nanotechnology (ELNANO), April 16-18 2019, Kyiv, Ukraine;

- 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD), October 22-24 2019, Kyiv, Ukraine;

- 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), October 2019, Kyiv, Ukraine.

Публікації. За результатами досліджень опубліковано 16 наукових праць, у тому числі 5 статей у наукових фахових виданнях України, які включені до міжнародних наукометричних баз, 1 свідоцтво про авторське право на комп'ютерну програму, 10 тез доповідей в збірках матеріалів конференцій.

Структура та об'єм дисертаційної роботи. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та двох додатків. Загальний обсяг роботи становить 161 сторінок, у тому числі 116 сторінок основного тексту, 79 рисунків та списку використаних джерел із 77 найменувань.

РОЗДІЛ 1. СУЧАСНИЙ СТАН АКУСТИЧНОЇ ЕКСПЕРТИЗИ КОМУНІКАЦІЙНИХ КАНАЛІВ

1.1. Спотворення мовних і музичних сигналів в комунікаційних каналах

Відомо [6], що основними параметрами аналого-цифрового акустичного тракту є: смуга частот, рівномірність амплітудної частотної характеристики, лінійність фазової частотної характеристики, динамічний діапазон, частота дискретизації, кількість рівнів квантування.

Низку робіт [1-22] присвячено дослідженням впливу цих параметрів на якість звучання музики та мовного сигналу.

Мовні й музичні сигнали неминуче спотворюються, при перетворенні різними електроакустичними пристроями, такими як мікрофони, фільтри, мікшери, компресори, еквайзери, підсилювачі та гучномовці. Крім того, сигнали спотворюються в цифрових пристроях, таких як АЦП та ЦАП, системи кодування та декодування. Зауважимо втім, що в ряді технічних пристроїв мову навмисно спотворюють, домагаючись у такий незвичний спосіб підвищення її розбірливості. Прикладом таких пристроїв є слухові апарати, де застосовується звичайне або центральне кліпування, а також спеціальні системи відновлення сигналів, записаних на тлі інтенсивних перешкод. Наведені приклади свідчать про те, що, якість і розбірливість мови – поняття тісно пов'язані, але не тотожні.

Сигнали в комунікаційних каналах практично завжди спотворюються через дії шуму, реверберації (рис. 1.1), дефектів частотних характеристик трактів передачі, помилок кодування. Тому до складу каналів передачі мовних сигналів зазвичай включають системи корекції, що пригнічують шум, реверберацію, а також компенсують дію інших несприятливих чинників. І хоча користь від систем корекції може бути досить відчутною, абсолютно точно відновити спотворені сигнали практично неможливо. Більш того, системи корекції, зазвичай, самі вносять специфічні спотворення в мовний сигнал. У зв'язку з цим

досить важливо вміти оцінювати якість не тільки систем корекції, але й трактів передачі в цілому [6].

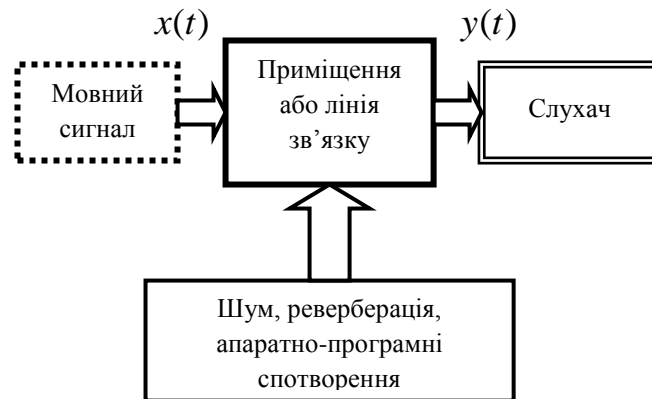


Рис. 1.1. Вплив факторів, що спотворюють мовний сигнал [6].

Останнім часом в трактах мовної комунікації все ширше використовуються системи автоматичного розпізнавання мови (АРМ), вельми чутливі не тільки до різниці умов навчання та, власне, розпізнавання, але й до дії шуму та реверберації. У зв'язку з цим виникає завдання забезпечення високої якості мовних сигналів та, як наслідок, забезпечення стійкості систем АРМ до дії зазначених факторів.

Приміщення також можна розглядати як різновид комунікаційних каналів. При проходженні через повітря та інші середовища звукові хвилі не тільки поступово втрачають свою потужність, але й змінюють свій спектральний склад через такі фізичні явища як залежність ступеня поглинання звуку середовищем від частоти коливань, дисперсія швидкості звуку.

Слід зазначити, що в комунікаційних каналах практично завжди корисні сигнали спотворюються шумовими перешкодами, як природного походження, так і техногенного характеру. Розвиток систем мобільного зв'язку, розширення практики проведення відеоконференцій пов'язані з необхідністю не тільки враховувати наявність шумових перешкод, але й пригнічувати ці завади. Слід, однак, враховувати, що системи пригнічення шуму роблять свій внесок у спотворення сигналів.

У закритих приміщеннях, де звукові хвилі багато разів відбиваються від стін та предметів інтер'єру, спотворення акустичних сигналів носять специфічний характер. Ранні відбиття дещо підсилюють мовний або музичний сигнали, тоді як пізні відбиття («пізня реверберація») впливають на слух як певний різновид шумової завади.

Найпростішим засобом забезпечення високої якості мовного сигналу в точці його прийому є розширення смуги пропускання каналу зв'язку. Так, наприклад, якщо для аналогового телефонного зв'язку прийнятною вважалася смуга частот до 3,1 кГц, сучасні цифрові телефонні мережі (стандарт ITU-T) орієнтовані на три смуги частот: вузьку (300 Гц – 3,4 кГц), широку (50 Гц – 7 кГц) та надшироку (50 Гц – 14 кГц). Більш того, поява в 2008 році кодека G.719, призначеного для цифрового зв'язку в смузі частот 20 Гц – 20 кГц, поставило на порядок денний впровадження «повночастотного» або «повнодіапазонного» (fullband) зв'язку, що дозволяє якісно передавати музику.

Крім обмеженості смуги частот, на якість переданого акустичного сигналу впливає багато інших факторів:

- кліпування
- шум і реверберація приміщення;
- помилки систем придушення шуму й реверберації;
- нерівномірність амплітудної частотної характеристики каналу передачі;
- нелінійність фазової характеристики каналу передачі;
- затримка сигналу;
- електронне відлуння;
- артефакти, створювані пригнічувачем відлуння;
- помилки кодування-декодування сигналів.

Що стосується навчальних приміщень, їхні акустичні характеристики істотно впливають на якість сприйняття навчального матеріалу. При цьому,

поряд з якістю мовного сигналу важливу роль відіграє такий показник як розбірливість мови.

У період з 1999 по 2009 рр. Д. Бредлі, Х. Сато і В. Янгом [25-28] було отримано кількісні оцінки ступеня впливу шуму і реверберації на розбірливість мови в навчальних приміщеннях. При цьому експериментально показано, що шумова перешкода є значно небезпечнішою за реверберацію, по-перше, через близькість основних джерел шуму, якими є учні, що сидять поруч. Ситуація ускладнюється тим, що мовний шум має високу маскувальну властивість через подібність спектрів завади і сигналу. Разом з тим, час реверберації в навчальних приміщеннях рідко перевищує 0,9 с, через що дифузна частина звукових відбиттів, іменована «пізньою реверберацією», має порівняно низьку інтенсивність і тому має слабкий маскувальний ефект.

Іншим важливим результатом, представленим в роботах [25-28], є кількісна оцінка залежності розбірливості мови від віку слухачів. Так, показано, що для досягнення словесної розбірливості 95% для школярів 6 років відношення сигнал-шум має бути не нижче 16 дБ, для школярів 11 років поріг може бути знижений до 9 дБ, а для студентів – до 1 дБ. Цей результат дозволяє зробити висновок, що рівень освіти людини тісно пов'язаний з його здатністю до правильного сприйняття інформації, переданої в усних повідомленнях на тлі перешкод.

1.2. Суб'єктивний і об'єктивний підходи до оцінювання якості мови і музики

1.2.1. Суб'єктивні міри якості мови і музики

Роль суб'єктивних методів є високою з огляду на їх еталонний характер: результати об'єктивної акустичної експертизи, як правило, в кінцевому рахунку повинні бути співвіднесені з результатами суб'єктивного оцінювання.

У [1] виділено три види тестування шляхом прослуховування, з виставленням оцінок за шкалами:

- абсолютній (Absolute Category Rating – ACR);
- ступеня деградації (Degradation Category Rating – DCR);
- порівняння якості (Comparison Category Rating – CCR).

Тестування систем зв'язку найчастіше здійснюють методом ACR, згідно з яким якість сигналу, що прослуховується, оцінюється тільки за результатами прослуховування тестового сигналу – порівняння із «зразковим ідеальним» сигналом при цьому не проводиться. Відповідну п'ятибальну шкалу оцінювання (табл. 1.1) називають MOS-Listening Quality Score (MOS_{LQS}) [1].

Таблиця 1.1. Шкала MOS_{LQS} [1]

Якість мови	Бали
Відмінно	5
Добре	4
Посередньо	3
Погано	2
Дуже погано	1

При оцінюванні методом DCR сигнал, що прослуховується, порівнюють з еталонним сигналом високої якості – відповідну шкалу (табл. 1.2) називають Degradation Mean Opinion Score (DMOS) [1].

Таблиця 1.2. Шкала DMOS [1]

Деградація якості мови	Бали
Нечутна	5
Чутна, але не дратує	4
Трохи дратує	3
Дратує	2
Дуже дратує	1

При суб'єктивному оцінюванні якості музичних сигналів використовується ряд мір якості, при цьому найбільш розповсюдженою також є деградація середнього показника думки (DMOS) [9] (табл. 1.3).

Таблиця 1.3. Оцінювання якості музичного сигналу за шкалою DMOS

DMOS	Значення
5	Якість музичного сигналу, що прослуховується, схожа з якістю еталонного, завади не чутно
4	Якість музичного сигналу дещо гірша за якість неспотвореного еталону, завади мінімальні
3	Якість музичного сигналу відчутно гірша у порівнянні з еталоном, чутно завади;
2	Якість музичного сигналу значно гірша у порівнянні з еталоном, завади майже на рівні корисного сигналу
1	Якість музичного сигналу значно гірша у порівнянні з еталоном, рівень завади дорівнює або є більшим за рівень корисного сигналу.

1.2.2. Об'єктивні міри якості мови і музики

Принциповою перевагою об'єктивного (інструментального) підходу до оцінки якості акустичних сигналів є його швидкість та відносна дешевизна [1-4]. Найбільш достовірними є так звані «перцептуальні» міри якості, що ґрунтуються на моделях слухової системи людини. Оскільки алгоритми та обчислювальні програми таких мір якості є досить складними, зрозумілим є бажання використовувати більш прості для обчислення міри якості. При цьому, однак, з особливою гостротою виникає питання достовірності використовуваних мір якості.

З відомих на сьогоднішній день об'єктивних показників якості мовних сигналів, кількість яких сягає кількох тисяч [2], в даній дисертаційній роботі

обмежимося розглядом найбільш вживаних у практичних застосуваннях показників: це сегментне відношення сигнал-шум (signal-to-noise ratio – SSNR), логарифмічно-спектральні спотворення (logarithmic-spectral distortion – LSD), зважене в частотній області сегментне відношення сигнал-шум (frequency-weighted segmental signal-to-noise ratio – fwSSNR), барк-спектральні спотворення (bark-spectral distortion – BSD) і перцепційна оцінка якості мови (perceptual evaluation of speech quality – PESQ) [1, 2]. Для широкосмугових сигналів (таких як музичні) замість міри PESQ прийнято використовувати перцепційну оцінку якості звуку (perceptual evaluation of audio quality – PEAQ). Обґрунтовуючи такий вибір, зазначимо, що перші два показники – SSNR і LSD – є вельми привабливими в силу простоти обчислень, тоді як інші три показники – fwSSNR, BSD та PESQ – дозволяють врахувати, з різним ступенем точності, особливості слухової системи людини. Аналітичний опис згаданих вище показників SSNR, LSD і BSD:

$$SSNR = \frac{1}{M} \sum_{m=1}^M 10 \lg \left[\frac{\sum_{n=R(m-1)+1}^{Rm} x^2(n, m)}{\sum_{n=R(m-1)+1}^{Rm} [x(n, m) - y(n, m)]^2} \right] \quad (1.1)$$

$$LSD = \frac{2}{KL} \sum_l \sum_{k=0}^{\frac{K}{2}-1} |G\{X(l, k)\} - G\{Y(l, k)\}|, \quad (1.2)$$

де $G\{X(l, k)\} = \max\{20 \lg(|X(l, k)|), \delta\}$, $\delta = \max_{l, k} \{20 \lg(|X(l, k)|)\} - 50$,

$$BSD = \frac{\sum_l \sum_{k=0}^{\frac{K}{2}-1} [B\{X(l, k)\} - B\{Y(l, k)\}]^2}{\sum_l \sum_{k=0}^{\frac{K}{2}-1} [B\{X(l, k)\}]^2}, \quad (1.3)$$

де $x(n, m)$ і $y(n, m)$ – m -та вибірка n -го фрейму вхідного та вихідного сигналів фільтра $x(n)$ та $y(n)$, відповідно; $X(l, k)$ та $Y(l, k)$ – амплітудні спектри l -го фрейму

сигналів $x(n)$ та $y(n)$, відповідно; $B\{X(l,k)\}$ и $B\{Y(l,k)\}$ – барк-спектри l -го фрейму сигналів $x(n)$ та $y(n)$, відповідно.

До групи показників частотної області також належить частотно-зважене сегментне відношення сигнал-шум (fwSSNR):

$$fwSSNR = \frac{10}{M} \sum_{m=1}^M \frac{\sum_{k=1}^K W(k,m) \lg \frac{|X(k,m)|^2}{(|X(k,m)| - |Y(k,m)|)^2}}{\sum_{k=1}^K W(k,m)}, \quad (1.4)$$

де k – номер критичної смуги; K – кількість критичних смуг; $|X(k,m)|$ – амплітудний спектр чистого сигналу; $W(j,m)$ – вагові коефіцієнти.

Найпривабливішими в плані простоти обчислень є такі міри як сегментне відношення сигнал-шум SSNR і лог-спектральні спотворення LSD. Практичне використання цих мір показало, що вони мають ряд особливостей, які, на жаль, недостатньо висвітлені в літературі. Ця прогалина частково заповнюється в роботах [4-7], а в даній дисертаційній роботі пропонується як уточнення попередніх, так і низка нових результатів.

Узагальнена структурна схема алгоритму PESQ наведена на рис. 1.2[1], з неї випливає, що на першому етапі обчислень використовується перцептуальна модель слухової системи. Відповідно до цієї моделі, сигнал трансформується з часової області в спектральну область за допомогою перетворення Фур'є, а потім шкали частот і рівнів отриманого спектра трансформуються в шкали барків і гучності. Тобто, як і у випадку з показником BSD, обчислюються барк-спектри сегментів мови. Показаний на рис. 3.6 блок вирівнювання покликаний нівелювати часові затримки сегментів спотвореного сигналу, зумовлені різними технічними причинами. Однак ясно, що наявність даного блоку не є принципово важливим чинником, оскільки такий блок повинен бути присутнім в алгоритмі обчислень будь-якого об'єктивного показника якості мови.

Принциповою відмінністю показника PESQ від показника BSD є наявність когнітивної моделі, за допомогою якої враховуються такі ефекти як асиметрія,

відмінність ваг, а також компенсуються лінійні спотворення. Ефект асиметрії проявляється в тому, що спотворення на виході кодеків відчуються сильніше тоді, коли кодеки спотворюють існуючий частотно-часовий компонент, ніж тоді, коли кодеки пропускають цей компонент. Ефект відмінності ваг полягає в тому, що спотворення, які сталися під час активної мови, відчуються як більш сильні в порівнянні з спотвореннями, що сталися під час пауз. Нарешті, під компенсацією лінійних спотворень мається на увазі часткове усунення тих спотворень, які були внесені системою при фільтрації мовних сигналів. У результаті показник PESQ виявляється здатним оцінювати якість мовних сигналів, спотворених кодеками, помилкою передачі бітів, шумовою завадою, фільтрацією, затримками, в тому числі часовими затримками.

Міру PEAQ застосовують для оцінки якості музичного сигналу аудіопристрою, що тестується, або аудіокодеку. PEAQ порівнює вхідний і вихідний аудіосигнали цього пристрою та видає показник якості, який представляє якість вихідного сигналу, а отже, і якість пристрою. Порівняння двох сигналів враховує лише перцептивні відмінності, тоді як непомітними спотвореннями нехтують. PEAQ включає моделі слухової системи, засновані на швидкому перетворенні Фур'є, а також на гребінці фільтрів. Вихідні значення моделей базуються частково на концепції маскувального порогу і частково на порівнянні внутрішніх уявлень (також відомих як порівняння в кохлеарному домені). Крім того, міра PEAQ також дає вихідні значення на основі порівняння лінійних спектрів, які не обробляються вушною моделлю.

Вихідні значення відображаються в єдиний показник якості – об'єктивну різницю якості (ODG – objective difference grade) за допомогою штучної нейронної мережі з одним прихованим шаром. ODG показує виміряну базову якість звуку досліджуваного сигналу за постійною шкалою від -4 (порушення, що дуже дратує) до 0 (непомітні порушення). Крім ODG, модель видає додаткове значення якості – індекс спотворення (DI). Узагальнену схему алгоритму PEAQ наведено на рис. 1.3 [2].

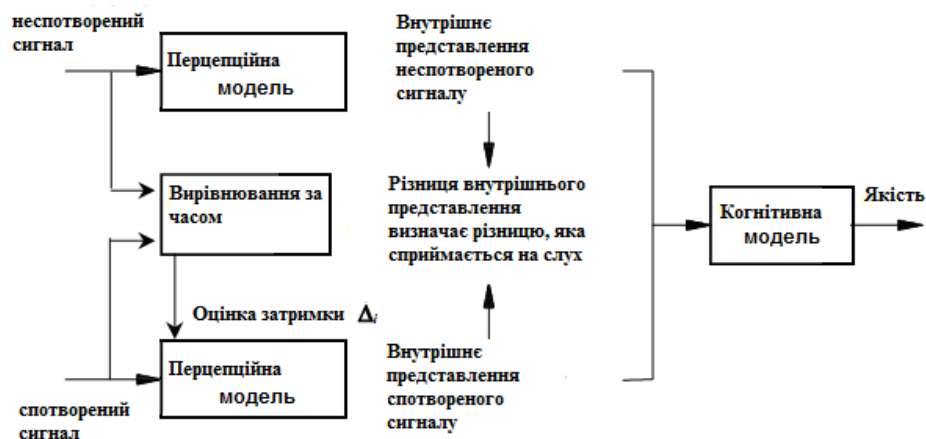


Рис. 1.2. Узагальнена схема алгоритму PESQ [1]



Рис. 1.3. Узагальнена схема алгоритму оцінювання міри PEAQ [2]

Суттєвим недоліком показників PESQ та PEAQ є високий рівень складності програмного забезпечення. Окрім того, міра PESQ на даний час вважається морально застарілою, замість неї останніми роками використовується міра POLQA, програмне забезпечення якої, на жаль, відсутнє у відкритому доступі.

У роботі [5] розглянуто ряд досліджень факторів, що впливають на якість звуку, в основному для прикладів джазової і класичної музики. Дані були

отримані з використанням оцінок окремих подразників або за допомогою методу парних порівнянь. Для учасників з нормальним слухом, найвищі оцінки якості звуку були отримані, коли ширина смуги відтворення була широка (від 55 Гц до 16 кГц) і флуктуації частотної характеристики були невеликими (менше ± 5 дБ).

Об'єктивне й суб'єктивне оцінювання якості мовних сигналів з обмеженою смугою частот було проведено в роботі [7], де використано такі об'єктивні міри якості мовних сигналів як SSNR, LSD, fwSNR, BSD та PESQ, а також побудовано карти відповідності результатів об'єктивного й суб'єктивного оцінювання якості мовленнєвих сигналів. Важливим є висновок про те, що отримані карти відповідності є більш корисними в прикладному плані, ніж коефіцієнти кореляції, оскільки припускають можливість нелінійної залежності між об'єктивними й суб'єктивними оцінками якості мовленнєвого сигналу та виконують функцію калібрувального коефіцієнту. Проте одержані в [7] результати не є вичерпними, оскільки стосуються тільки мовленнєвих сигналів. Для усунення вказаного недоліку необхідно провести подібні дослідження також і для музичних сигналів.

Результати оцінювання впливу нелінійності фазової частотної характеристики тракту на якість мовних та музичних сигналів наведено в роботах [8] та [9], де показано, що для слухової системи людини прийнятними є фазові спотворення сигналів, якщо максимальна різниця групових часів затримки в області високих і низьких частот не перевищує 70 мс для музики та 50 мс для мови. При цьому об'єктивні показники SSNR, LSD, BSD і PESQ в цілому адекватно реагують на ступінь фазових спотворень музичного сигналу. Оскільки отримані в роботах [8] та [9] результати слід вважати попередніми з огляду на невеликий обсяг даних, підданих статистичній обробці, для уточнення отриманих результатів бажано усунути цей недолік.

Окрім розглянутих вище випадків лінійного спотворення акустичних сигналів, в трактах передачі акустичної інформації можливі також нелінійні

спотворення, розповсюдженим різновидом яких є обмеження миттєвих значень сигналу (кліпування).

Загальною особливістю наукових праць, присвячених задачі виявлення кліпування, є необхідність попереднього оцінювання щільності розподілу ймовірності (probability density function – PDF) аналізованого сигналу [12]-[19].

Так, зокрема, у патенті США [12] розглянуто чотири варіанти реалізації запропонованого способу, які відрізняються вимогами до об'єму апіорної інформації або способом оцінювання ступеня кліпування.

У російському патенті [13] також в якості міри ступеня кліпування запропоновано використовувати параметри PDF у вигляді дисперсії, середньоквадратичного відхилення, а також такі параметри як середньовипрямлене значення та середня кількість викидів. Ненормований характер перерахованих параметрів є серйозним недоліком, що обмежує практичне використання даного способу.

Зазначеного недоліку позбавлена пропозиція використовувати відношення сигнал-шум в ролі міри ступеня кліпування [14]. При цьому під сигналом мається на увазі неспотворене миттєве значення аудіосигналу, а під шумом мається на увазі миттєві значення аудіосигналу, які зазнали спотворення. Оскільки миттєві значення такого шуму невідомі, його потужність пропонується оцінювати по екстрапольованим хвостам PDF аналізованого сигналу. Очевидним недоліком даного способу є його громіздкість, зумовлена як необхідністю оптимізації параметрів алгоритму екстраполяції, так і наявністю самого алгоритму екстраполяції.

Евристичний підхід до вибору міри ступеня кліпування, запропонований в [15], базується на тому факті, що кліпування призводить до появи сплесків на кордонах PDF. Тому мірою ступеня кліпування запропоновано вважати дистанцію між вершинами цих сплесків і центральним піком щільності ймовірностей, нормовану за діапазоном можливих значень кліпованого сигналу.

Разом з тим, в [15] самокритично зазначено, що запропонований спосіб є недостатньо точним для оцінювання ступеня кліпування.

У способі виявлення кліпування, описаному в [16], вирішується завдання виявлення кліпування, для чого запропоновано попередньо побудувати досить грубу, на 20 бінів, гістограму абсолютних значень аналізованого сигналу. Для прийняття рішення про наявність кліпування, запропоновано використовувати значення гістограми для найвищих значень сигналу. Головним недоліком даного способу є відсутність етапу попереднього оцінювання ступеня кліпування.

Вибір мел-частотних кепстральних коефіцієнтів в якості класифікаційних ознак [19] для виявлення кліпування пояснюється особливістю розв'язуваної задачі, що складається в віддаленій (за допомогою телефонного або мережевого зв'язку) діагностиці хвороби Паркінсона за особливостями вимови голосного звуку /а/. Вочевидь, основним недоліком даного способу є його порівняно висока обчислювальна складність.

На жаль, у жодній зі згаданих робіт не розглядався можливий претендент на роль міри ступеня кліпування у вигляді кумулятного коефіцієнту четвертого порядку, який іноді (для стислості) називають коефіцієнтом ексцесу [20-23]. Частково цю прогалину заповнено в [20], де на прикладі аналізу мовних сигналів продемонстровані принципова можливість і перспективність використання коефіцієнта ексцесу не тільки як міри ступеня кліпування, але і як міри якості кліпованих сигналів. Даний напрямок досліджень видається дуже перспективним з огляду на простоту та швидкість алгоритму обробки, що дозволяє уникнути попереднього обчислення щільності розподілу ймовірності сигналу.

1.3. Суб'єктивний і об'єктивний підходи до оцінювання розбірливості мови

Розбірливість мови також в певному сенсі характеризує якість акустичних сигналів. Проте, на відміну від якості мови, для оцінювання розбірливості мови використовують шкалу відношень, а не порядкову шкалу, оскільки під

розбірливістю мови розуміють частку елементів мови (склади, слова, фрази), правильно сприйнятих людиною.

На сьогодні існує два підходи до вимірювання розбірливості мови: суб'єктивний, при якому мовна та слухова системи людини є основними елементами вимірювальної системи, та об'єктивний (інструментальний). Хоча суб'єктивний підхід є досить обтяжливим в сенсі значних затрат людських, часових та фінансових ресурсів, його безсумнівна та першорядна цінність полягає в можливості одержати еталонні результати, необхідні в подальшому для калібрування систем об'єктивного вимірювання [1].

1.3.1. Суб'єктивний підхід до оцінювання розбірливості мови

Оскільки на сьогодні Україна не має відповідного національного стандарту суб'єктивного оцінювання розбірливості мови, інженери змушені користуватися стандартом колишнього СРСР ГОСТ 16600-72 [23]. В Росії замість цього вже давно використовують стандарт 50840-95 [24], який осучаснено наданням можливості використовувати ПК та спеціальне програмне забезпечення для автоматизації окремих етапів суб'єктивної акустичної експертизи.

Важливою особливістю частини стандарту 50840-95 [24], що стосується такої автоматизації, є застосування таблиць неповних звукосполучень та процедури дописування букв, відсутніх в звукосполученнях, за допомогою клавіатури комп'ютера. Перевагою такого підходу, як вказано в [24, 25], є значне скорочення часу навчання артикуляційної бригади. Однак, на наш погляд, недоліком такого підходу є його суттєва відмінність від «класичного» суб'єктивного підходу, де слухачі не дописують звукосполучення, а записують їх цілком. Недоліком дописування складів є те, що слухачеві фактично підказується частина складу, що прослуховується.

Кроком до створення національного стандарту України можна вважати роботу [26], де запропоновано десять артикуляційних таблиць українських слів. Результати тестування цих таблиць шляхом суб'єктивних вимірювань розбірливості мови на тлі білого та забарвленого шумів представлено в [27]. На жаль, при цьому не вказано режим використання генератора забарвленого шуму ANG 2200 [28], а за представленим в [27] графіком оцінки спектру забарвленого шуму важко збагнути, який тип забарвлених шумів було використано. Насторожує й те, що представлені в [27] результати одержано шляхом прослуховування тестових сигналів через звукові колонки комп'ютера. При цьому не вказано відстань слухачів до колонок, а також не вказано характеристики приміщення, в якому відбувалося прослуховування, хоча, як відомо, ранні відбиття звуку в приміщеннях можуть суттєво, до 6-9 дБ, підвищувати відношення сигнал-шум в місці прослуховування [28].

1.3.2. Об'єктивний підхід до оцінювання розбірливості мови

Об'єктивні методи оцінювання розбірливості мови, такі як формантний та модуляційний (рис. 1.4), є на сьогодні найбільше детально дослідженими та розповсюдженими, хоча емпіричні методи, які полягають в обчисленні показників C50 та %ALcons, є також досить популярними.

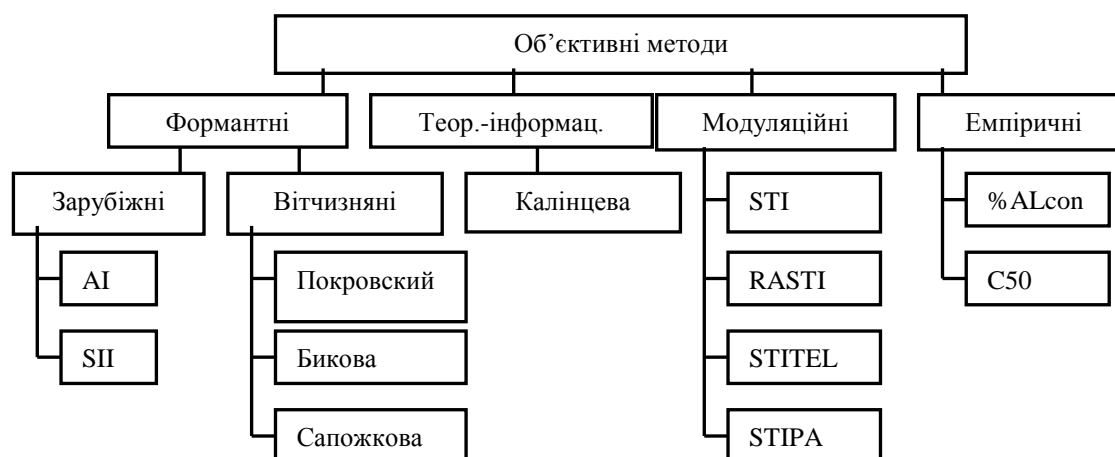


Рис. 1.4. Класифікація об'єктивних методів оцінки розбірливості мови [6]

Принциповою перевагою об'єктивних методів є швидкість та відносна дешевизна вимірювань.

Особливу роль при оцінюванні розбірливості мови виконують тестові сигнали. Якщо для суб'єктивних методів це елементи реальної мови – склади, слова, фрази, то для об'єктивних методів – це спеціальні тестові сигнали. При використанні формантного методу у місці, де, зазвичай, розташовується диктор, розміщують гучномовець, за допомогою якого випромінюють тестовий сигнал у вигляді відрізка стаціонарного шуму, спектр якого подібний до довготривалого спектру мови. У місці, де передбачається можливе знаходження слухача, розміщують мікрофон, вихідний сигнал якого записують та піддають обробці для оцінювання розбірливості мови.

Словесну розбірливість W обчислюють через формантну розбірливість A [71]:

$$W = \begin{cases} 1.54 \cdot A^{0.25} [1 - \exp(-11 \cdot A)], & A < 0.15; \\ 1 - \exp\left(\frac{11 \cdot A}{1 + 0.7 \cdot A}\right), & A \geq 0.15. \end{cases} \quad (1.5)$$

$$A = \sum_{k=1}^K p_k \cdot P(E'_k), \quad (1.6)$$

де $P_k(E'_k)$ – коефіцієнт сприйняття мови; $E_k \approx 10 \lg \frac{D_{sk}}{D_{nk}}$ – відношення сигнал-шум в k -й смузі частот; p_k – ймовірність перебування формант в k -й смузі частот з центральними частотами f_{0k} та граничними частотами f_{1k} та f_{2k} :

$$p_k = F(f_{2k}) - F(f_{1k}), \quad (1.7)$$

$$F(f) = \begin{cases} 2.57 \cdot 10^{-8} \cdot f^{2.4}, & 100 < f \leq 400 \text{ Гц}; \\ 1 - 1.074 \cdot \exp(-10^{-4} \cdot f^{1.18}), & 400 < f \leq 10000 \text{ Гц}. \end{cases} \quad (1.8)$$

При оцінюванні розбірливості мови модуляційним методом або тісно пов'язаним із ним формантно-модуляційним методом [6], випромінюють тестовий сигнал $x(t)$ у вигляді нестационарного випадкового процесу (НВП), що є результатом модуляції, за гармонічним законом з частотою модуляції F ,

стаціонарного випадкового процесу $\xi(t)$ із дисперсією D_ξ та зі спектром потужності, що повторює форму довготривалого спектра мови:

$$x(t) = \xi(t)\sqrt{f(t)}, \quad (1.9)$$

$$f(t) = 1 + \cos 2\pi Ft. \quad (1.10)$$

Дисперсія $D_x(t)$ процесу $x(t)$ при цьому має вигляд:

$$D_x(t) = \langle x^2(t) \rangle - \langle x(t) \rangle^2 = D_\xi(1 + \cos 2\pi Ft), \quad (1.11)$$

де $\langle \cdot \rangle$ – символ математичного очікування. Вочевидь, коефіцієнт модуляції m дисперсії $D_x(t)$ дорівнює одиниці.

У місці, де передбачається можливе знаходження слухача, спостерігається сигнал $y(t)$:

$$y(t) = \int_{-\infty}^{\infty} h(v)x(t-v)dv + n(t). \quad (1.12)$$

де $n(t)$ – шум; $h(v)$ – імпульсна характеристика лінії зв'язку або приміщення.

Дисперсія $D_y(t)$ сигналу $y(t)$ також є модульованою за гармонічним законом з такою ж частотою F , однак, внаслідок дії ревербераційної і шумової завад, коефіцієнт модуляції m дисперсії $D_y(t)$ виявляється меншим за одиницю. Наприклад, у випадку, коли превалює шумова завада, у співвідношенні (1.12) можна прийняти $h(v) \approx \delta(v)$, де $\delta(v)$ – дельта-функція Дірака, внаслідок чого:

$$D_y(t) = D_\xi(1 + \cos 2\pi Ft) + D_n, \quad (1.13)$$

$$m = \frac{D_\xi}{D_\xi + D_n}. \quad (1.14)$$

Зі співвідношення (1.14) випливає вираз для відношення сигнал-шум:

$$SNR = 10 \lg \frac{D_\xi}{D_n} = 10 \lg \frac{m}{1-m}. \quad (1.15)$$

При спільній дії шумової та ревербераційної перешкод аналітичний вираз для коефіцієнта модуляції m дисперсії $D_y(t)$ виявляється більш складним у порівнянні з (1.14) і залежить не тільки від дисперсії шуму D_n , але і від стандартного часу реверберації T_{60} , і навіть від частоти модуляції F . Але принциповим є те, що співвідношення (1.15) надалі можна використовувати для обчислення «еквівалентного» відношення сигнал-шум в k -й смузі частот за наявності не тільки шумової завади, але й реверберації, що є перевагою модуляційного (та формантно-модуляційного) методу, порівняно із формантним методом.

До емпіричних методів оцінювання розбірливості мови належать %ALcons та C50. Показник %ALcons базується виключно на врахуванні відсотку приголосних, які слухачі не можуть зрозуміти в приміщенні, оскільки голосні не є нагально потрібними для розуміння повідомлення. Показник розбірливості мови C50 є відношенням величин звукової енергії ранніх і пізніх відбить звуку у приміщенні [6]. Незважаючи на недоліки, зумовлені недостатньо високою точністю одержуваних результатів, емпіричні методи використовуються і донині, що пояснюється простотою й швидкістю відповідних обчислень [6].

Хоча якість та розбірливість мовних сигналів не є тотожними поняттями, проте ідея замінити оцінювання розбірливості мови шляхом оцінювання якості мовних сигналів видається раціональною, по-перше, з огляду на певний зв'язок між цими поняттями, а по-друге, враховуючи складність оцінювання розбірливості мови, порівняно із відносною простотою оцінювання якості мовних сигналів. Нарешті, така заміна може вважатися доцільною, якщо дозволяє спростити акустичну експертизу без шкоди для достовірності отриманих результатів.

1.4. Багатомікрофонні масиви в комунікаційних каналах

На даний момент можна виділити три основні напрямки застосування просторово-часової обробки:

- бінауральний запис та відтворення;

- запис та відтворення акустичного поля;
- багатозональне відтворення.

Перший напрямок – бінауральний запис та відтворення – це такий запис, який призначений для створення враження просторового звучання при відтворенні за допомогою навушників або за допомогою двох гучномовців (стереосистеми). При такому підході вдається передати відчуття, що джерело звуку знаходиться правіше й лівіше. Другий напрямок – це використання великої кількості мікрофонів та гучномовців, при якому вдається створити набагато більш повне враження про об'ємний звук в деякій області простору. Третій напрямок – ситуація, коли в приміщенні створюються локальні зони озвучування незалежними сигналами (у великих офісах, у виставкових приміщеннях, в особистому транспорті і т.п.), включаючи створення зон тиші. Для вирішення цього завдання необхідні масиви гучномовців, вихідні сигнали яких управляються за спеціальними алгоритмами. Технологію «звукових об'єктів» називають також «форматно-агностическим підходом» (format agnostic approach), альтернативним традиційному багатоканальному підходу. (Форматно-агностичний підхід – це підрядковий переклад; по суті, мова йде про форматно-незалежний підхід). Справа в тому, що при традиційному багатоканальному підході формати, взагалі кажучи, можуть бути несумісними: наприклад, стерео-формат не може бути переведений в формат 5.1 і 7.1, планарні формати не можуть бути переведені в 3D-формати. Технологія звукових об'єктів дозволяє піти від цієї проблеми несумісності форматів – це і є сутністю форматно-незалежного підходу [46-65]. У [57] вказується на 2 способи запису акустичних сигналів: мікрофони розташовані близько до джерел звуку або запис проводиться віддаленими спрямованими системами. У першому випадку легко створити звуковий об'єкт, додаючи метадані, що описують місце розташування об'єкта. При цьому, однак, втрачається притаманна об'єктам інформація про їх просторові властивості. Цього недоліку позбавлені віддалені спрямовані мікрофонні системи.

Формування характеристики спрямованості або просторова фільтрація є одним з найпростіших засобів для розрізнення сигналів на основі фізичного розташування джерел. Оскільки музика є широкосмуговим сигналом, що охоплює інколи більше чотирьох октав, традиційний метод формування характеристики спрямованості, такий як метод сум та затримок (DS – delay and sum) [43] вузькосмугового сигналу, може видатися неспроможним для отримання якісного звучання. Сигнал на виході масиву мікрофонів, сформований за методом DS описується відношенням:

$$y(n, \theta_i) = \sum_{m=1}^M x_m(n - \Delta_m(\theta_i)), \quad (1.16)$$

де Δ_m затримка, розраховано відповідно напрямку надходження сигналу θ_i .

Найбільшим недоліком методу формування характеристики спрямованості методом DS є те, що розмір масиву визначається найнижчою частотою роботи. Габарит масиву, що має працювати в діапазоні музичних сигналів від 20 Гц до 20 кГц, має сягати декілька метрів. Тому видається перспективним клас широкосмугових формувачів характеристики спрямованості, що слугують для отримання вузької просторової характеристики, постійної в широкому діапазоні частот. Вузькість діаграми спрямованості та її незалежність від частоти забезпечуються шляхом використання в частотній області вагових коефіцієнтів для відліків спектрів сигналів на виходах мікрофонів. Після підсумовування зважених таким чином сигналів, вихідний сигнал отримують за допомогою оберненого перетворення Фур'є. Алгоритм обробки в таких масивах є більш складним, порівняно з (1.16):

$$y(t, \theta_i) = \sum_{m=1}^M w_m(\theta_i, \theta_i) x_m(t - \Delta_m(\theta_i)), \quad (1.17)$$

де Δ_m затримка, розрахована відповідно напрямку надходження сигналу θ_i , w_m – вагові коефіцієнти, які залежать від частоти та напрямку θ_i (у випадку DS антени вони дорівнюють 1).

У роботі [57] виконано порівняння 4 методів:

- класичний метод просторової фільтрації DS (delay-sum);
- суперспрямовані решітки (SDA);
- метод MDVR (minimum variance distortionless response);
- узагальнення методу MDVR у вигляді методу LCMV (linearly constrained minimum variance).

Метод SDA дозволяє мінімізувати ізотропний акустичний шум у всіх напрямках, окрім цільового. Вагові коефіцієнти для методу SDA визначаються співвідношенням [44]:

$$w(\theta_l) = \frac{(R_{vv} + \varepsilon I)^{-1} a(\theta_l)}{a(\theta_l)^H (R_{vv} + \varepsilon I)^{-1} a(\theta_l)}, \quad (1.18)$$

де $R_{vv} = \frac{1}{L} \sum_{l=1}^L a(\theta_l) a^H(\theta_l)$, I – одинична матриця розміром $M \times M$, ε – ваговий коефіцієнт, який керує підсиленням білого шуму відносно спрямованості, θ_l – l -й кут напрямку.

Алгоритм MVDR визначається співвідношенням:

$$\min_w w_m^H R_{xx} w \cdot w^H a(\theta_l) = 1 \quad (1.19)$$

де $R_{xx} = E[xx^H]$ – матриця просторової кореляції, θ_l – напрямок надходження сигналу, $E[\cdot]$ – оператор математичного очікування.

Алгоритм LCMV визначається співвідношенням:

$$\min_w w_m^H R_{xx} w \cdot w^H a(\theta_l) = 1; w^H a(\theta_i) = 0 \quad (1.20)$$

де θ_i – фіксований напрямок надходження шуму,

Зі співвідношення (1.20) отримано рівняння для функції вагових коефіцієнтів методу LCMV [57]:

$$w_m(\theta_t, \theta_t) = (R_{xx} + \varepsilon I)^{-1} C [C^H (R_{xx} + \varepsilon I)^{-1} C]^{-1} C^H w \quad (1.21)$$

де I – одинична матриця розміром $M \times M$, ε – параметр регуляризації, що залежить від частоти, C – матриця обмежень.

Детально обґрунтування методів MVDR і LCMV наведені в роботі [73].

Зазначимо, що спільною особливістю методів SDA, MVDR та LCMV є незалежність характеристики спрямованості від частоти сигналу. Перевагою методу LCMV є гнучкість при застосованні в ситуаціях виділення корисного звуку з декількох джерел та придушення реверберації приміщення [57].

Хоча в [44-56] та [58-65] виконано дослідження можливостей просторової обробки широкосмугових акустичних сигналів за допомогою мікрофонних решіток, однак суб'єктивну та об'єктивну оцінку якості сигналів на виходах мікрофонних масивів не проведено. Лише частково питання оцінки якості сигналів, оброблених мікрофонними решітками розкрито в роботі [57], де представлені наступні міри для оцінювання якості мікрофонних масивів:

- параметри, що визначаються ваговими коефіцієнтами спрямованої системи – індекс спрямованості (DI – directivity index), чутливість до білого шуму (WNG – white noise gain) та акустичний контраст (AC – acoustic contrast) між напрямками на джерело завади;
- перцепційні оцінки – це такі об'єктивні оцінки, як оцінка якості розрізнення джерел звуку (perceptual evaluation for audio source separation – PEASS), а також оцінка якості мови (PESQ) та її аналог у вигляді оцінки якості музики (PEAQ);
- оцінки відношення сигнал/шум (SNR), відношення сигнал/спотворення (signal-to-distortion ratio – SDR), відношення сигнал/артефакти (signal-to-artifact ratio – SAR), відношення сигнал/шум для сусідніх джерел звуку, які є завадами (contamination by interfering sources – SIR).

Досить несподіваним в роботі [57] є те, що, за результатами об'єктивного оцінювання якості мовних сигналів, класичний метод DS виявився кращим за інші методи (рис. 1.5).

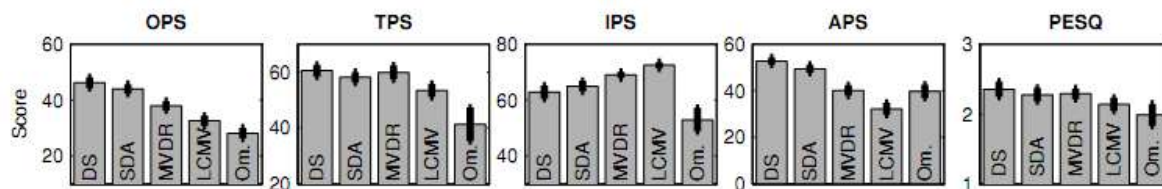


Рис. 1.5. Оцінки якості мовного сигналу для різних видів формування характеристики спрямованості в масиві мікрофонів [57]

Винятком є міра IPS, що є цілком зрозумілим, оскільки звичайна мікрофонна решітка забезпечує найгіршу просторову фільтрацію джерел.

На жаль, в [57] не наведено результатів суб'єктивного оцінювання якості музичних сигналів на виході алгоритмів просторово-часової обробки, що суттєво знижує цінність отриманих даних з огляду на еталонний характер результатів суб'єктивного оцінювання якості акустичних сигналів.

Висновки до розділу 1

1. Аналіз існуючих методів оцінки якості акустичних сигналів свідчить про те, що перевагою суб'єктивного оцінювання якості сигналів є еталонний характер отриманих результатів, оскільки якість звукових сигналів оцінюється кінцевими користувачами. Суттєвим недоліком суб'єктивного оцінювання є високі вимоги до фінансових, часових та людських ресурсів. Такого недоліку позбавлені об'єктивні (інструментальні) методи й системи оцінювання.

2. Головною проблемою при застосуванні методів об'єктивного оцінювання є різноманітність причин спотворення акустичних сигналів. Іншою проблемою є надзвичайно велика кількість об'єктивних мір якості сигналів, що значно ускладнює вибір потрібної міри. На жаль, залишаються не до кінця дослідженими потенційні можливості простих для обчислень об'єктивних мір,

таких, зокрема, як сегментне відношення сигнал-шум та лог-спектральні спотворення.

3. Нагальною проблемою є отримання еталонних результатів суб'єктивного оцінювання якості та розбірливості мови. Проте отримання достовірних еталонних результатів суб'єктивного оцінювання потребує обробки великих обсягів звукового матеріалу, що призводить до необхідності автоматизувати процедуру артикуляційних випробувань. На жаль, системи автоматизації артикуляційних випробувань, що використовувалися до останнього часу, є надто складними та дорогими. Тому актуальним є питання створення простих, недорогих та надійних систем автоматизації оцінювання якості та розбірливості мови.

4. Важливою особливістю мовних сигналів є те, що якість та розбірливість мови не є тотожними поняттями: якісний сигнал не завжди є розбірливим, а розбірливий сигнал не завжди є якісним. Оскільки оцінювання розбірливості мови об'єктивними методами є набагато складнішим, порівняно з оцінюванням якості мови, виникає питання про пошук умов, за яких можливе використання простих мір якості замість складних мір розбірливості мови. На жаль, це важливе питання й досі залишається недостатньо дослідженим.

5. Недостатньо вивченим на теперішній час є також питання оцінювання якості музичних сигналів. Розглянуті дотепер проблеми стосуються переважно оцінювання параметрів музичних сигналів, а питання об'єктивного оцінювання якості спотворених музичних сигналів в них майже не висвітлено. Складність таких мір та високі вимоги до обчислювальних ресурсів є їхнім суттєвим недоліком, що мотивує до пошуків більш простих і водночас достатньо надійних технічних рішень.

6. У задачах просторово-часової обробки акустичних сигналів за допомогою мікрофонних решіток на сьогоднішній день недостатньо дослідженими є питання суб'єктивного оцінювання якості сигналів, а також питання доцільності використання складних алгоритмів просторово-часової обробки замість класичного алгоритму сум та затримок.

РОЗДІЛ 2. ОЦІНЮВАННЯ ЯКОСТІ МОВИ І МУЗИКИ ПРИ ЛІНІЙНИХ І НЕЛІНІЙНИХ СПОТВОРЕННЯХ

2.1. Суб'єктивне оцінювання якості мовних та музичних сигналів

Суб'єктивні методи можна розділити на дві групи, виходячи з необхідної вхідної інформації [2]:

- Неінтрузивні методи полягають у відображенні передбачуваного значення оцінки за шкалою MOS
- Інтрузивні методи полягають у порівнянні спотвореного сигналу з еталонним та відображенні передбачуваного значення оцінки за шкалою DMOS.

У даній дисертаційній роботі перевагу надано інтрузивним суб'єктивним методам, оскільки, при експертизі комунікаційних каналів практично завжди є наявності еталонний сигнал.

З плином часу методи суб'єктивного оцінювання якості мовних та музичних сигналів все більше витісняються об'єктивними (інструментальними) методами, як більш швидкими і дешевими. На жаль, у значній кількості досліджень [1-6] достовірність об'єктивних мір якості оцінюється з використанням коефіцієнту кореляції між результатами об'єктивної та суб'єктивної експертизи. Такий підхід є недостатньо коректним, оскільки при цьому фактично не допускається можливість нелінійної залежності між результатами об'єктивної та суб'єктивної експертизи. Тому в підрозділах 2.2-2.6 значну увагу приділено не тільки об'єктивним мірам якості мовних та музичних сигналів, але й такому важливому питанню як оцінювання достовірності об'єктивних мір якості за допомогою карт відповідності результатів об'єктивної та суб'єктивної експертизи. Такі карти є важливими ще й для практичних застосувань, оскільки дозволяють здійснити калібрування об'єктивних мір якості мовних та музичних сигналів.

2.2. Об'єктивне оцінювання якості мовних та музичних сигналів

2.2.1. Залежність результатів обчислень міри SSNR від похибки вирівнювання сигналів за часом

Істотну залежність результатів обчислень міри SSNR від похибки вирівнювання сигналів за часом сигналів, що порівнюються, наведено в [3], хоча кількісна оцінка ступеня такого впливу не наведена. У [72] цю прогалину заповнено, і для найпростішої моделі гармонійного сигналу таку кількісну оцінку отримано. Дійсно, вважаючи:

$$x(t) = \cos 2\pi f_0 t, \quad y(t) = \cos 2\pi f_0 (t - \tau), \quad (2.1)$$

приходимо до простого співвідношення:

$$SSNR = SNR = 10 \lg \frac{\int_0^T x^2(t) dt}{\int_0^T [x(t) - y(t)]^2 dt} = 10 \lg \frac{0,5}{1 - \cos 2\pi f_0 \tau}, \quad (2.2)$$

де f_0 – частота сигналу, τ – похибка вирівнювання в часі, T – інтервал спостереження. Конкретні числові значення $SSNR(\tau)$ для різних співвідношень F_s/f_0 , де F_s – частота дискретизації, наведено в табл. 2.1.

Таблиця 2.1. Значення $SSNR(\tau)$

τF_s F_s/f_0	1	0,5
22,05	10,93 дБ	16,93 дБ
2	-6,02 дБ	-3,01 дБ

Оскільки на практиці значення SSNR, зазвичай, обмежують зверху величиною 35 дБ, неважко дійти висновку, що похибка вирівнювання

гармонічних сигналів всього в один період дискретизації ($\tau F_s = 1$) призводить до похибки оцінювання SSNR, близької $35 - 10,93 \approx 24$ дБ для сукупності значень $f_0 = 1$ кГц та $F_s = 22050$ Гц. Для $f_0 \approx 11$ кГц, при тій самій частоті дискретизації, похибка сягає $35 - (-6,02) \approx 41$ дБ. Настільки швидке зростання похибки легко пояснити, якщо врахувати, що при $f_0 \approx 11$ кГц та $F_s = 22050$ Гц на одному періоді гармонічного сигналу розміщується всього дві вибірки цього сигналу. Тому простий і досить очевидний спосіб боротьби з зазначеним явищем полягає в інтерполяції сигналів $x(n)$ та $y(n)$, спрямованій на поліпшення «промальовування» високочастотних компонентів [72].

У табл. 2.1 також наведено значення SSNR для помилки вирівнювання, яка дорівнює половині відстані між окремими вибірками ($\tau F_s = 0,5$). Така помилка має місце, якщо $y(n)$ є результатом фільтрації сигналу $x(n)$ нерекурсивним фільтром парного порядку. Звідси випливає висновок про необхідність контролювати порядок нерекурсивних фільтрів, який повинен бути непарним [72].

Ефективність даних рекомендацій продемонструємо на прикладі оцінки якості мовних сигналів, обмежених за смугою частот Δf [72]. На рис. 2.1а показана залежність $SSNR(\Delta f)$, отримана без урахування наведених вище рекомендацій, а на рис. 2.1 б) – з урахуванням таких. У даному прикладі фігурують залежності $SSNR(\Delta f)$, отримані усередненням 8 оцінок для чоловічого та жіночого мовлення. Як бачимо, виявилось достатнім підвищити F_s (шляхом інтерполяції) з 22050 Гц (рис. 2.1а) до 44100 Гц (рис. 2.1б), одночасно забезпечуючи непарний порядок нерекурсивних НЧ фільтрів. На рис. 2.1в наведено результати суб'єктивного оцінювання якості цих же мовних сигналів [6] з використанням шкали Degradation Mean Opinion Score (DMOS) [3] (в експерименті брало участь 17 слухачів, юнаків та дівчат віком 21-22 років).

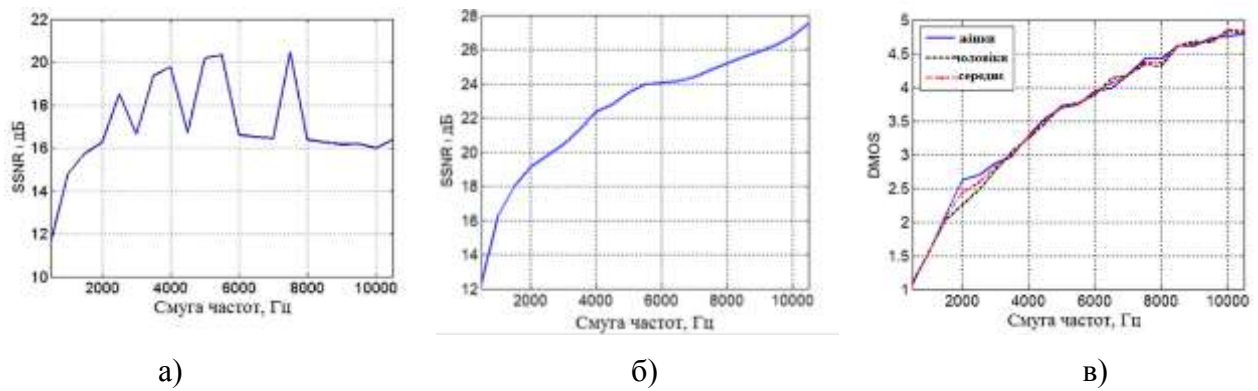


Рис. 2.1. Залежності $SSNR(\Delta f)$ до врахування особливостей SSNR (а) і після врахування таких особливостей (б), а також суб'єктивна оцінка якості мовних сигналів (в) [7], [72]

2.2.2. Особливості оцінки міри LSD

Міра якості мовних сигналів, що називається лог-спектральними спотвореннями (Log-Spectral Distortion – LSD), описується співвідношенням:

$$LSD_{(p)} = \left(\frac{1}{F} \int_0^F |LX(f) - LY(f)|^p df \right)^{1/p}, \quad (2.3)$$

$$LX(f) = 20 \lg |X(f)|, \quad LY(f) = 20 \lg |Y(f)|,$$

де $|X(f)|$ и $|Y(f)|$ – спектри амплітуд еталонного процесу $x(t)$ та спотвореного процесу $y(t)$, відповідно, F – смуга частот, яка аналізується, $p=1,2,\dots,\infty$ – параметр, що визначає різновид міри. У [7, 8] показано існування зв'язку міри LSD_2 з кепстральними мірами. Незважаючи на важливість цих результатів, слід констатувати, що властивості міри LSD та її оцінок вивчені недостатньо. Цим пояснюється порівняно нечасте її використання при експериментальних дослідженнях.

У ряді досліджень міра LSD_1 показала себе як «суб'єктивно значуща» [2-6], тобто суб'єктивно більшим спотворенням сигналу відповідали більші значення міри LSD_1 . Однак при оцінюванні якості мовних сигналів, обмежених смугою частот Δf , виявилося, що залежність $LSD_1(\Delta f)$ містить локальні екстремуми (рис. 2.2а, які не зникають зі збільшенням часу аналізу.

Для вивчення даного феномена доцільно використовувати дві моделі сигналів: білий шум і вузькосмуговий стаціонарний випадковий процес [74]. Обґрунтуванням такого вибору моделей може слугувати той факт, що в мовному сигналі чергуються широкосмугові (приголосні) і вузькосмугові (голосні) звуки.

Представлені результати досліджень – моделі білого шуму показали, що залежність $LSD_1(\Delta f)$ є немонотонною та описується співвідношенням:

$$LSD(kF) = LSD^*(k) = 10[k|\lg k| + |\lg k + 2\lg M| \cdot (1-k)], \quad (2.4)$$

де $k = \Delta f / F$ ($0 \leq k \leq 1$) – нормована смуга пропускання НЧ фільтра з прямокутною АЧХ, M – кількість фреймів. Слід зауважити, що графіки залежності (2.4), показані на рис. 2.2б, отримані для нормованих по стандартному відхиленню сигналів. За відсутності такого нормування отримуємо лінійну залежність $LSD^*(k) = 50 \cdot (1-k)$, $0 \leq k \leq 1$, графік якої представлений на рис. 2.2в.

Переходячи до моделі вузькосмугового сигналу, представимо її у вигляді відгуку рекурсивного смугового фільтра другого порядку на вплив у вигляді дискретного білого шуму x_n :

$$y_n = a_0 x_n - b_1 y_{n-1} - b_2 y_{n-2}, \quad (2.5)$$

де $a_0 = 1 + b_1 + b_2$, $b_1 = -2\rho \cos(\theta)$, $b_2 = \rho^2$, $\theta = 2\pi(f_c/F_s)[1 - 0.25(F_s/f_w)^2]^{1/2}$, $\rho = \exp(-\pi f_c/F_s)$, f_c та f_w – центральна частота і ширина смуги пропускання фільтра, відповідно; F_s – частота дискретизації. Форма АЧХ фільтру для $f_c = 700$ Гц та $f_w = 50$ Гц показана на рис. 2.3а. Цю АЧХ можна розглядати також як модель форманти мовного сигналу.

Графіки залежності $LSD(\Delta f)$ (при нормуванні сигналів, що зіставляються) для формант з частотами від 100 до 700 Гц наведено на рис. 2.3б. Як бачимо, на

додаток до «гострого» локального мінімуму і наступного за ним «плавного» локального максимуму, поблизу частоти форманти f_c з'являється ще один невеликий локальний максимум. За відсутності нормування сигналів залежності $LSD(\Delta f)$ монотонно спадають (рис. 2.3в).

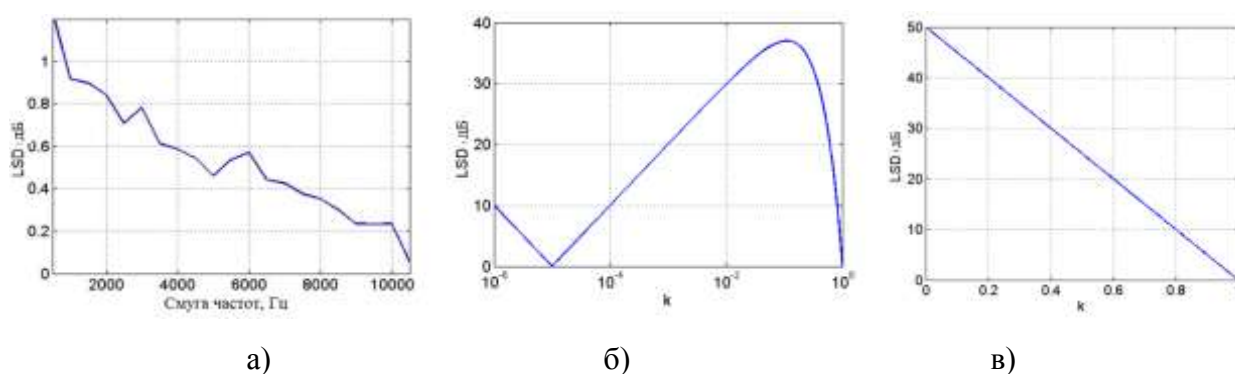


Рис. 2.2. Усереднена по 8 дикторам залежність $LSD_1(\Delta f)$ (а) та залежність $LSD^*(k)$ для детермінованої версії моделі білого шуму: при нормуванні по стандартному відхиленню (б) і за відсутності нормування (в)

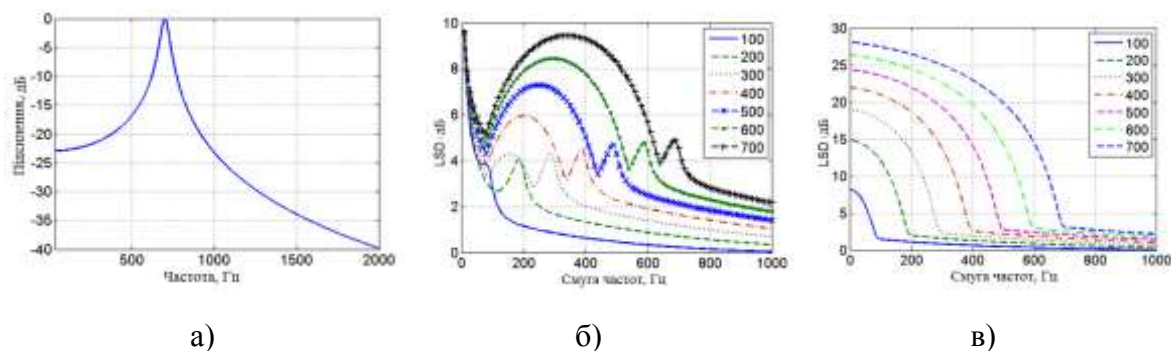


Рис. 2.3. Частотна характеристика фільтру для $f_c=700$ Гц, $f_w=50$ Гц (а) та залежність $LSD(\Delta f)$ для моделі вузькосмугового сигналу: при нормуванні по стандартному відхиленню (б) та за відсутності такого (в)

Результати дослідження залежності $LSD(\Delta f)$ для реальних мовних сигналів наведені на рис. 2.4. При цьому, щоб позбутися артефактів, обумовлених формою частотних характеристик НЧ фільтрів, обмеження смуги частот сигналів

виконувалось за допомогою НЧ фільтрів з ідеально прямокутними АЧХ (фільтрація в частотній області). Відповідні оцінки $LSD(\Delta f)$ наведені на рис. 2.4 для сигналів, нормованих за стандартним відхиленням.

Як бачимо, порушення монотонності залежності $LSD(\Delta f)$ спостерігається лише при сильних спотвореннях мовних сигналів (смуга частот 500 Гц і нижче), що можна пояснити близькістю центру формант до зазначеної смуги частот. Як показано далі в п. 2.3, для музичних сигналів залежності $LSD(\Delta f)$ не є монотонними в значно ширшому діапазоні частот, що можна пояснити більшою складністю спектрів сигналів.

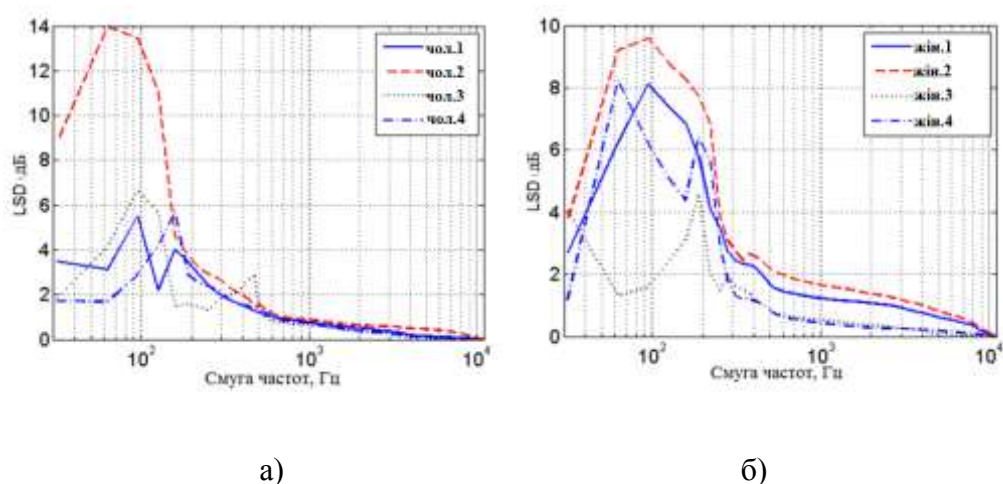


Рис. 2.4. Оцінки $LSD(\Delta f)$ для мови: чоловіки (а), жінки (б)

2.3. Об'єктивне оцінювання якості музичних сигналів, обмежених за смугою частот

У роботі [7] шляхом обчислення коефіцієнтів кореляції та побудови карт відповідності зіставлено результати об'єктивного та суб'єктивного оцінювання якості мовних сигналів, спотворених обмеженням смуги частот. При цьому передбачалося, що мовні сигнали займають смугу частот, яка не перевищує 11 кГц. Однак для якісного відтворення музичних сигналів необхідно

використовувати значно ширший частотний діапазон, що сягає 20 кГц, що й було зроблено при отриманні представлених нижче результатів [69], [75].

2.3.1. Пілотні дослідження якості музичних сигналів, обмежених за смугою частот

Під час пілотних експериментальних досліджень [75] якості музичних сигналів, обмежених за смугою частот, використано фрагменти музичних творів тривалістю 16 секунд, у категоріях «популярної» та «класичної» музики. Формат файлів – WAVE, з частотою дискретизації 44,1 кГц. Обладнання – ноутбук із зовнішнім аудіо інтерфейсом, що підтримує програвання сигналу з частотою дискретизації 44,1 кГц та навушники.

Нижче наведені назви файлів і відповідних композицій.

Зразки класичної музики:

- Й.-С. Бах – Концерт Мі мажор (Алегро) # 1;
- Й.-С. Бах – Концерт Ре мажор (Алегро) # 1;
- Й.-С. Бах- Концерт Ре мажор (Алегро) # 3;
- Й.-С. Бах – Концерт Ре мінор (Алегро) # 3.

Зразки естрадної музики:

- 3 Doors Down – Here without you;
- Hoobastank – Reason;
- One Republic – Stop and Stay;
- You me at six – Lived a lie.

Спотворені фрагменти музичних сигналів формувалися шляхом почергового пропущення їх через нерекурсивні фільтри низької частоти (НЧ), синтезовані методом Ремеза [8]. Характеристики НЧ-фільтрів:

- частота зрізу змінювалася від 1 до 20 кГц з кроком 1 кГц;
- розмір перехідної зони становив 5% від смуги пропускання;

- нерівномірність АЧХ в смузі пропускання становила 1 дБ;
- загасання в смузі придушення становило мінус 70 дБ.

Суб'єктивне оцінювання якості музичного сигналу здійснювалося 30-ю юнаками і дівчатами у віці 20-25 років, без вад слуху.

Усереднені результати суб'єктивного оцінювання (рис. 2.5, суцільна лінія), свідчать про практично монотонне зростання якості музичного сигналу зі збільшенням смуги частот. Штриховими лініями показані межі 95%-го довірчого інтервалу.

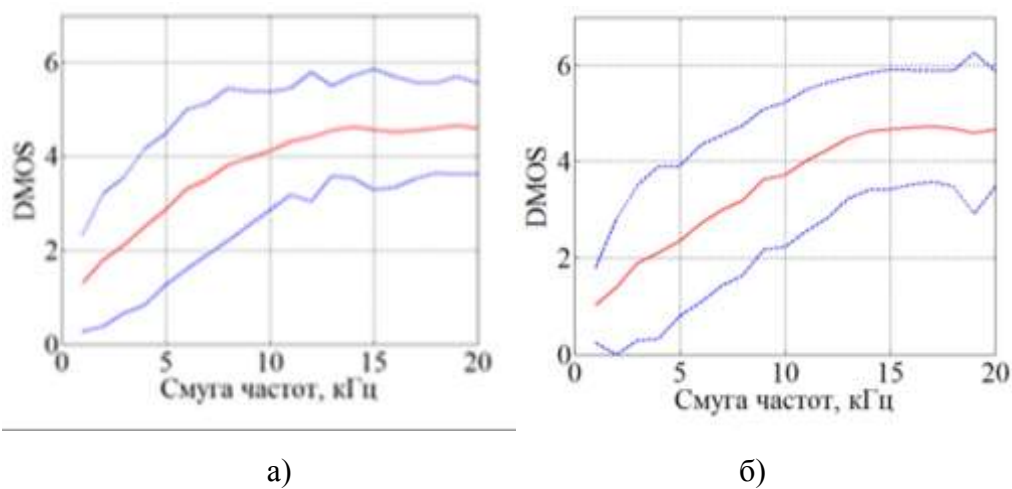
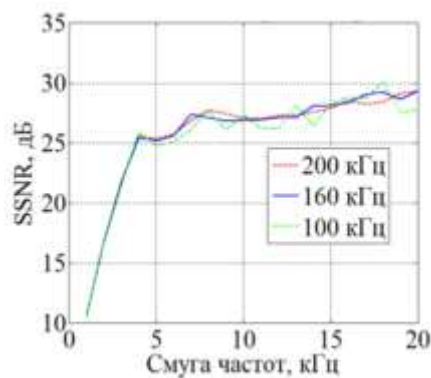


Рис. 2.5. Усереднені результати суб'єктивного оцінювання за шкалою DMOS класичної (а) та популярної (б) музики

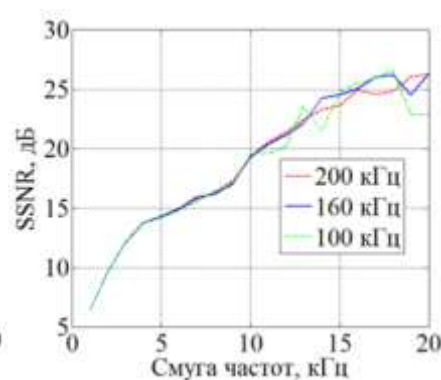
З наведених графіків видно, що для слуху людини достатньо смуги частот 14-15 кГц, в межах якої якість музичних композицій є найвищою, згідно шкали DMOS. Однак, слід зазначити значний розкид думок слухачів щодо якості матеріалу, який прослуховується. Дані результати добре узгоджуються з вимогами до смуги частот УКВ радіостанцій [6], що складає $\Delta f = 15$ кГц. Таким чином, отримані в даній дисертації результати суб'єктивного оцінювання є достовірними.

Результати об'єктивного оцінювання якості цих же музичних сигналів наведені на рис. 2.6. З огляду на зазначену в [7] чутливість міри SSNR до значення частоти дискретизації, в даній роботі дослідженню піддавалися як

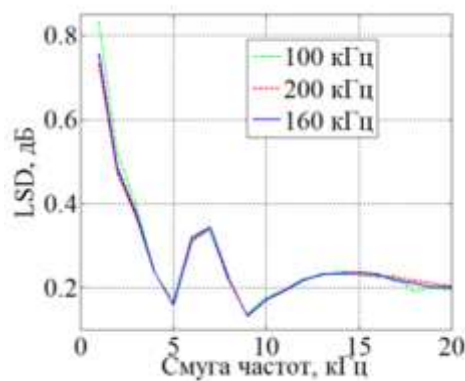
сигнали з вихідною частотою дискретизації 44,1 кГц, так і ці ж сигнали, піддані передискретизації до частот 100 кГц, 160 кГц і 200 кГц. Як впливає з наведених результатів, підвищення частоти дискретизації позитивно впливає на оцінки показника SSNR, дозволяючи усунути алогічні порушення монотонності залежності SSNR від смуги частот Δf . Цікавим виявився факт впливу процедури інтерполяції на точність вимірювань показника LSD: у випадку класичної музики (рис. 2.6в) монотонність залежності $LSD(\Delta f)$ порушується, при цьому сама функція $LSD(\Delta f)$ не залежить від частоти дискретизації. Інакше поводить себе залежність $LSD(\Delta f)$ у випадку популярної музики (рис. 2.6г), де функція $LSD(\Delta f)$ залежить від частоти дискретизації, причому порушення монотонності стає виразнішим із підвищенням частоти дискретизації. Що стосується оцінок показника BSD, чутливості до частоти дискретизації не виявлено.



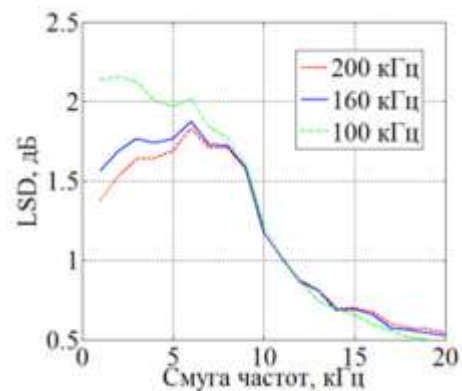
а)



б)



в)



г)

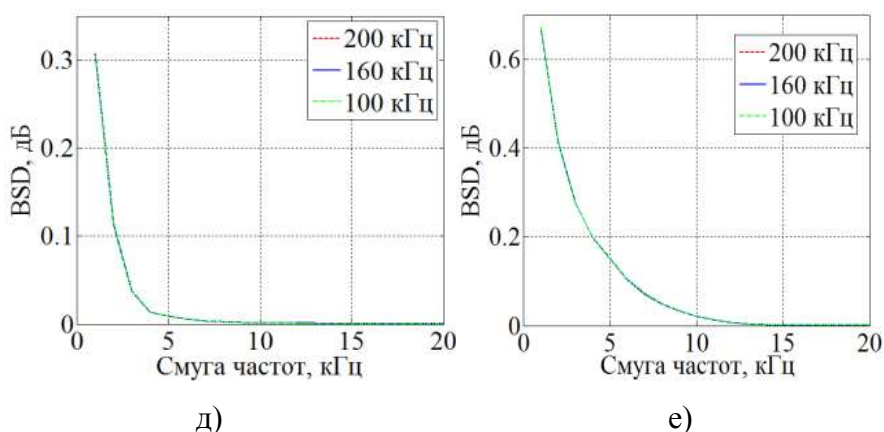


Рис. 2.6. Залежності SSNR(Δf), LSD(Δf) та BSD(Δf) для класичної (а, в, д) та популярної (б, г, е) музики

З графіків на рис. 2.6 видно, що всі розглянуті об'єктивні міри виявилися чутливими до спектральних особливостей класичної та естрадної музики. Проте поведінка графіків LSD(Δf) відрізняється вираженими порушеннями монотонності. Так, для класичної музики характерним є порушення монотонності залежності LSD(Δf) в інтервалі значень $\Delta f = 0,5-1,4$ кГц, а для естрадної музики – в інтервалі $\Delta f = 1-5$ кГц. При цьому, залежність LSD(Δf) для класичної музики практично не залежить від частоти дискретизації, тоді як у випадку популярної музики, форма LSD(Δf) істотно залежить від частоти дискретизації. Дослідження показали, що однією з причин порушення монотонності залежності LSD(Δf) є не ідеально прямокутна форма АЧХ використовуваних НЧ фільтрів. Дійсно, наведені на рис. 2.9 залежності, отримані при виконанні фільтрації в частотній області з АЧХ ідеальної прямокутної форми, є монотонними.

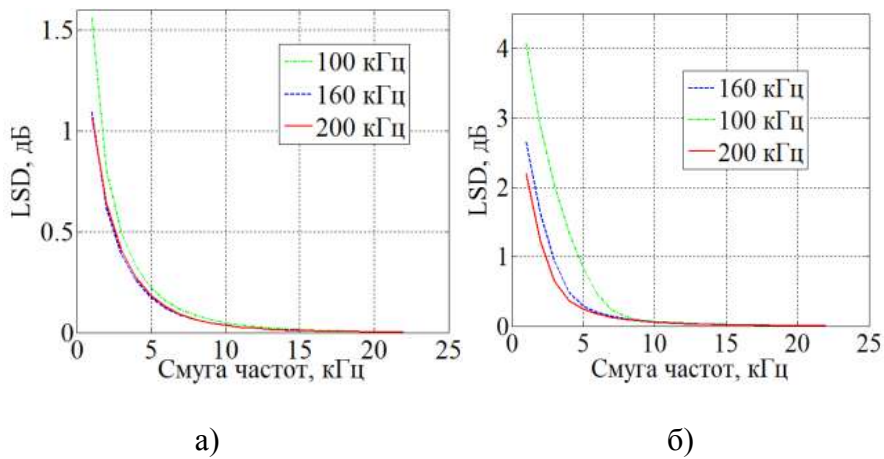


Рис. 2.7. Залежність LSD (Δf) для фільтрів з прямокутною АЧХ для класичної (а) та популярної (б) музики

Відзначимо, що при використанні НЧ фільтрів з прямокутними АЧХ значення $LSD(\Delta f)$ також виявилися залежними від вибору частоти дискретизації, як і при використанні «неідеальних» НЧ фільтрів.

Таким чином, аналіз якості музичних сигналів, обмежених за смугою частот, показав, що для отримання монотонної залежності $SSNR(\Delta f)$ частоту дискретизації слід збільшувати в більшому ступені, аніж це робилося для мовних сигналів, а саме – в 4-5 разів. Як бачимо, середнє значення оцінок DMOS монотонно і плавно підвищується до значення $\Delta f = 14$ кГц, після чого стабілізується на рівні 4,5-4,7. Графіки оцінок $SSNR(\Delta f)$ помітно відрізняються: до 4 кГц відбувається швидке зростання, а після 4 кГц швидкість росту помітно падає. Як і в разі мовних сигналів, можна говорити про досить хорошу узгодженість результатів об'єктивного і суб'єктивного оцінювання, що свідчить про дієвість запропонованих рекомендацій щодо необхідності інтерполяції сигналів перед оцінюванням показника $SSNR$.

Зауважимо, що хоча в пілотних дослідженнях при суб'єктивних випробуваннях брали участь 30 осіб, дану вибірку навряд чи можна вважати репрезентативною, тому у подальшому її обсяг слід збільшити. Крім того, досить обмеженою є вибірка музичних творів, що тестувалися. З огляду на перелічені

особливості, отримані результати слід розглядати як попередні. Таким чином, було прийнято рішення провести розширений експеримент зі збільшеними обсягами вибірок музичних творів, результати якого наведено нижче.

2.3.2. Уточнюючі дослідження якості музичних сигналів, обмежених за смугою частот

В уточнюючих дослідженнях [69] також виконувалося суб'єктивне тестування за методом DMOS. У суб'єктивному оцінюванні якості музичних сигналів брали участь 23 слухачі, юнаки та дівчата віком 19-24 років без явних вад слуху, при цьому використовувалися фрагменти музичних творів тривалістю від 15 до 30 секунд.

Дослідження показників SSNR, LSD, BSD, а також PEAQ здійснювалося в середовищі Matlab із застосуванням фрагментів музичних сигналів тривалістю 40 секунд.

У якості еталонних музичних сигналів було обрано 8 музичних творів, по 2 твори для кожного із 4-х жанрів: класична музика, поп-музика, джаз та рок-музика. Якість запису обраних творів відповідала стандарту Audio CD, тобто еталонні сигнали мали частоту дискретизації 44100 Гц та бітову глибину 16 біт.

Зразки поп-музики:

- LP – Lost on you;
- Karmin – Dance with you.

Зразки рок-музики:

- System of a Down – Question!
- Kings of Leon – Sex on fire.

Зразки джазової музики:

- Avishai Cohen – Nu Nu;
- Laco Tayfa – Atamaca.

Зразками класичної музики, що оцінювалися, були:

- Johannes Brahms – Hungarian Dance #5;
- П.І. Чайковський – «Пори року» / Серпень.

Спотворення музичних сигналів здійснювалося шляхом їх низькочастотної (НЧ) фільтрації нерекурсивними фільтрами Ремеза із такими ж характеристиками, як і в пілотних дослідженнях (п. 2.3.1).

На рис. 2.8 представлено усереднені результати суб'єктивного оцінювання якості музичних сигналів, де границі 95%-го довірчого інтервалу показано вертикальними відрізками. Як вже вказувалося, з розширенням смуги частот від 1 кГц до 14 кГц спостерігається монотонне й досить швидке зростання якості сигналу. Подальше розширення смуги частот практично не призводить до зростання якості музичних сигналів, що добре узгоджується з результатами [7].

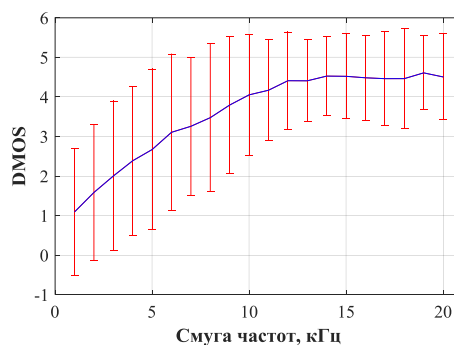


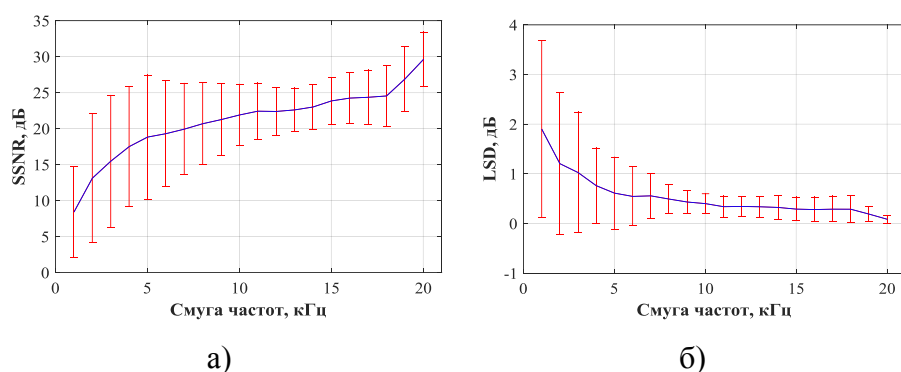
Рис. 2.8. Результати суб'єктивного оцінювання за шкалою DMOS

Стосовно об'єктивних показників, в цьому сенсі винятком є графік оцінки SSNR (рис. 2.9, а), де є досить різке пришвидшення росту якості сигналу при розширенні смуги частот понад 18 кГц. З одного боку, це пояснюється тим, що з розширенням смуги частот форма спотвореного сигналу наближається до форми еталонного сигналу, тому значення оцінки SSNR наближається до граничного значення +35 дБ. З іншого боку, можна говорити про наявність зміщення оцінки SSNR, спричиненого її чутливістю до значення частоти дискретизації. Як вказувалося в п. 2.2.1, це зміщення суттєво зменшується підвищенням в 2-4 рази, шляхом інтерполяції частоти дискретизації сигналів, що порівнюються.

У п.п. 2.2.2., 2.3.1 відмічено високу чутливість міри LSD до нерівномірності спектрів сигналів та до форми амплітудно-частотної характеристики (АЧХ) низькочастотного (НЧ) фільтру, наслідком чого є можливість порушення монотонності залежності $LSD(\Delta f)$. Графік оцінки LSD (рис. 2.9, б) має практично монотонний спадаючий характер, що пояснюється різноманітністю музичного матеріалу та усередненням оцінок LSD, внаслідок чого порушення монотонності залежності $LSD(\Delta f)$ практично зникають [58].

На графіку оцінки BSD (рис. 2.9в) покращення якості перестає бути помітним вже для $\Delta f > 10$ кГц. Що стосується поведінки оцінки PEAQ (рис. 2.9г), якість сигналу стабілізується для $\Delta f = 14$ кГц. Ледь помітне порушення монотонності спостерігаємо для смуг частот, менших за 3 кГц, проте цим порушенням можна знехтувати, зважаючи на те, що даний алгоритм розроблявся для сигналів із малим ступенем спотворення [7].

Оскільки достовірність об'єктивних мір якості сигналів часто визначають, розраховуючи значення коефіцієнтів кореляції між результатами об'єктивного і суб'єктивного оцінювання [7], доцільно розрахувати такі коефіцієнти кореляції для одержаних графіків рис. 2.8 та рис. 2.9 (табл. 2.2).



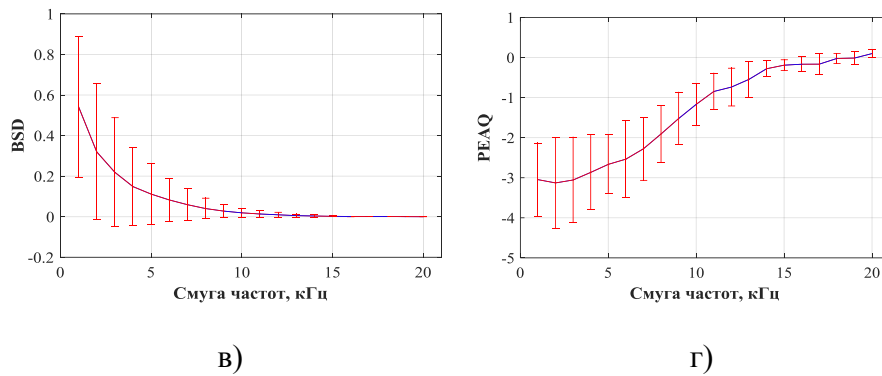


Рис. 2.9. Оцінки мір SSNR (а), LSD (б), BSD (в) та PEAQ (г)

Як бачимо, для всіх чотирьох мір одержано високі значення оцінок коефіцієнта кореляції. Проте слід зазначити, що такий підхід є не дуже коректним, оскільки між результатами об'єктивного і суб'єктивного оцінювання може існувати нелінійна залежність, яка не заважає встановленню взаємно однозначного зв'язку між результатами об'єктивного та суб'єктивного оцінювання. Тому для музичних сигналів, як і в роботі [7] для мовних сигналів, запропоновано замість обчислення коефіцієнтів кореляції будувати карти відповідності між значеннями об'єктивних та суб'єктивних оцінок якості.

Таблиця 2.2 Коефіцієнти кореляції оцінок якості

Об'єктивний показник	<i>SSNR</i>	<i>LSD</i>	<i>BSD</i>	<i>PEAQ</i>
Коефіцієнт кореляції	0.94	-0.93	-0.92	0.94

Безсумнівною перевагою карти відповідності перед коефіцієнтом кореляції є те, що така карта не тільки визначає характер залежності між суб'єктивними та об'єктивними оцінками, але і служить калібрувальним коефіцієнтом для перерахунку результатів об'єктивного оцінювання якості мовного сигналу в результати суб'єктивного оцінювання.

На рис. 2.10 представлено такі карти відповідності, побудовані з використанням графіків рис. 2.8 та рис. 2.9. Окрім експериментальних результатів (кружечки), на цих картах також представлено графіки лінійної та поліноміальної апроксимації.

У даному випадку оптимальним вважається такий мінімальний порядок поліному, при якому, з однієї сторони, досягається прийнятна для практичного застосування похибка апроксимації, а з іншої сторони, не порушується монотонність апроксимуючої функції.

Графіки рис. 2.10 побудовано з урахуванням такої пропозиції. Значення максимальних похибок апроксимації для перших п'яти порядків апроксимаційного поліному наведено в табл. 2.3 [69].

Таблиця 2.3 Максимальні похибки апроксимації [69]

Порядок поліному	Об'єктивний показник			
	<i>SSNR</i> , <i>дБ</i>	<i>LSD</i> , <i>дБ</i>	<i>BSD</i>	<i>PEAQ</i>
1	1.071	4.239	4.284	0.994
2	1.069	4.140	3.500	0.585
3	1.616	3.907	3.425	0.427
4	1.552	3.470	3.652	0.471
5	1.494	4.500	3.461	0.491

Як бачимо, карти відповідності для всіх чотирьох об'єктивних мір якості мають виражений нелінійний характер, незважаючи на високі значення коефіцієнту кореляції [69].

Досить неочікуваним може видатися той факт, що для карти відповідності PEAQ-DMOS (рис. 2.10 (г)), як і для інших розглянутих в даній роботі мір якості, також притаманний нелінійний характер. Здавалося б, оскільки міра PEAQ враховує особливості слухової системи людини, карта відповідності мала би

мати лінійний характер. Проте, якщо зважити на вже згадану вище особливість міри PEAQ, а саме, про її призначення для випадку малих спотворень сигналу, то на інтервалі $[-2,5...0]$ значень PEAQ дійсно спостерігаємо досить чітку лінійну залежність між PEAQ та DMOS. До речі, нелінійний характер інших карт відповідності в якійсь мірі також можна пояснити надмірними спотвореннями сигналу при малих значеннях смуги частот [69].

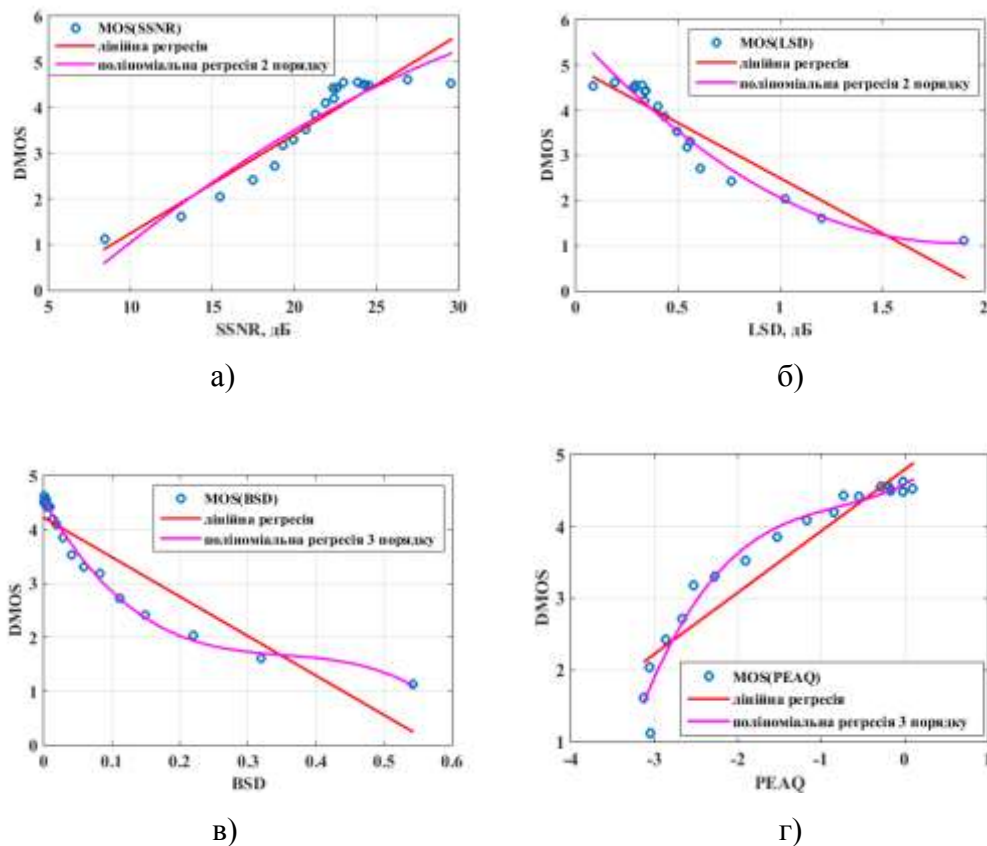


Рис. 2.10. Карти відповідності DMOS-SSNR (а), DMOS-LSD (б), DMOS-BSD (в) та DMOS-PEAQ (г)

Отримані суб'єктивні оцінки якості музичних сигналів від смуги частот каналу передачі дозволили підтвердити справедливості результатів попередніх досліджень, де вказано, що смуга частот 12-14 кГц є достатньою для того, аби музичний сигнал на слух практично не відрізнявся від еталонного сигналу. Проте слід враховувати, що зазначений результат є вірним «в середньому», тому не виключені випадки, коли окремі слухачі будуть вважати таку смугу частот недостатньо широкою для якісного відтворення музичних сигналів [69].

2.4. Оцінка ступеня впливу фазових спотворень на якість мови і музики.

Зазначимо, що одну із перших спроб визначення чутливості слухової системи людини до нелінійності фазової частотної характеристики $\theta(f)$ тракту передачі або, що те ж, до нерівномірності залежності групового часу затримки $\tau(f) = -\frac{1}{2\pi} \frac{d\theta(f)}{df}$, було зроблено в [76]. При цьому експериментально було показано, що спотворення сигналу на слух сприймаються, якщо нерівномірність $\tau(f)$ перевищує 1-3 мс. Для тренованого слуху цей поріг може бути знижений навіть до 0.4 мс. Однак згадані результати були отримані з використанням дуже коротких, протяжністю 25 мкс, імпульсів (або невеликих серій таких імпульсів). У [76] зазначено, що при використанні мовних і музичних сигналів фазові спотворення менш помітні. На жаль, відповідні порогові значення нерівномірності групової затримки не були визначені. Тим часом, для практичних цілей найбільший інтерес представляють саме мовні і музичні сигнали. Вперше спробу заповнити зазначену прогалину здійснено в роботах [8, 9], де показано, що для слухової системи людини прийнятними є фазові спотворення мовних і музичних сигналів, якщо максимальна різниця групових часів затримки тракту в області високих (в околиці 8 кГц) і низьких (в околиці 125 Гц) частот не перевищує 50-70 мс. У даній дисертаційній роботі наведено результати досліджень, проведених з метою уточнення результатів праць [8, 9]. При цьому, з метою підвищення достовірності результатів, до суб'єктивного оцінювання якості спотворених сигналів залучено значно більшу кількість слухачів. Крім того, вперше побудовано карти відповідності результатів об'єктивного та суб'єктивного оцінювання, що дає змогу виконати калібрування шкал об'єктивного оцінювання [66], [77].

У дослідженнях використовувалися сім октавних фільтрів, розрахованих методом Ремеза та аналогічних використаним в працях [8, 9]. Нерівномірність АЧХ в смузі пропускання становила 1 дБ, загасання в смузі затримання

становило мінус 60 дБ. Розміри перехідних зон змінювалися від 10%, для НЧ фільтрів, до 4%, для високочастотних фільтрів. Коефіцієнти нерекурсивних фільтрів розраховувалися в середовищі Matlab-Fdatool. Обраний спосіб розрахунку забезпечував мінімізацію порядку синтезованих фільтрів. Найбільш важливі параметри цих фільтрів наведені в табл. 3.1, де f_0 – центральна частота; Δf – смуга пропускання; n – порядок фільтру. Загальна смуга пропускання даної гребінки становить 90-11000 Гц як для мовних, так і для музичних сигналів [7].

Таблиця 2.2. Параметри гребінки октавних фільтрів

f_0 , Гц	125	250	500	1000	2000	4000	8000
Δf , Гц	90	180	355	710	1400	2800	5600
n	4353	2903	2177	1320	927	545	437

Незважаючи на те, що ФЧХ кожного з октавних фільтрів є лінійною, ФЧХ гребінки фільтрів, показаної на рис. 2.11а, є кусочно-нелінійною функцією частоти (рис. 2.12а). Груповий час затримки $\tau(f)$ при цьому є ступінчастою функцією частоти (рис. 2.12б). Пояснюється це тим, що ІХ гребінки фільтрів рис. 2.11а має несиметричну форму, оскільки протяжності ІХ кожного з смугових фільтрів різні.

Для управління ступенем і характером нелінійності функції $\theta(f)$ в кожному з каналів гребінки фільтрів програмним чином реалізовано затримку сигналу (рис. 2.11б).

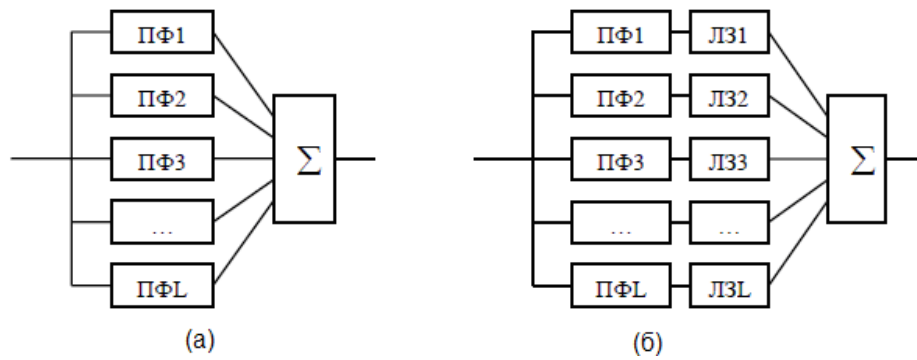


Рис. 2.11. Гребінка фільтрів з суматором (а) і з лініями затримки (б)

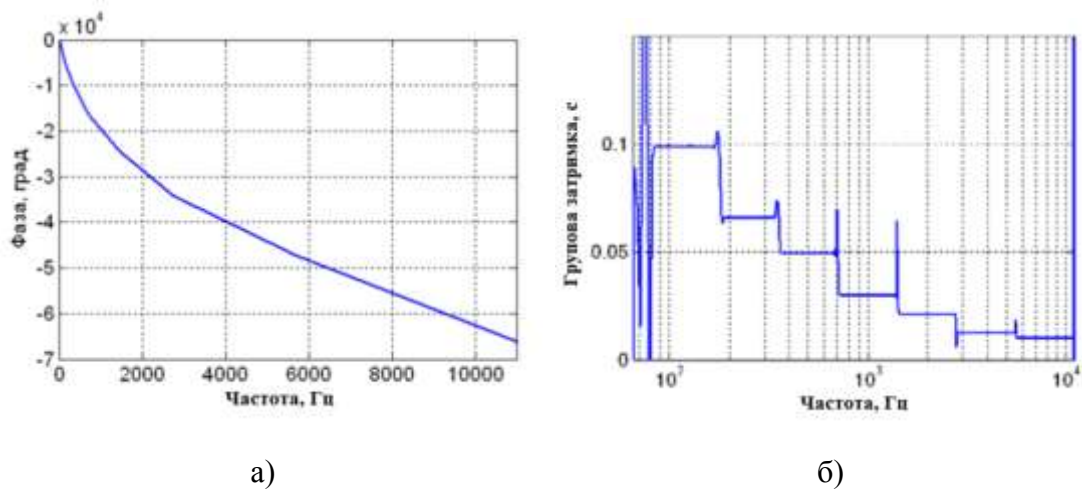


Рис. 2.12. Графіки ФЧХ (а) та $\tau(f)$ (б) для гребінки фільтрів із суматором

Показана на рис. 2.12б залежність $\tau(f)$ є східчасто-спадною, оскільки порядок фільтрів, що утворюють гребінку, спадає з ростом значення f_0 (див. табл. 2.2). Відповідну функцію $\theta(f)$ будемо позначати ФЧХ1. Ступінчасто-зростаючу залежність $\tau(f)$ можна отримати, затримуючи сигнали в частотних каналах гребінки фільтрів (рис. 2.11б). Функцію $\theta(f)$ для ступінчасто-зростаючої залежності $\tau(f)$ будемо позначати ФЧХ2.

Регулюючи величину затримки сигналів в каналах гребінки фільтрів, можна управляти не тільки типом залежності $\tau(f)$ (ФЧХ1 або ФЧХ2), але й ступенем нелінійності функції $\theta(f)$. У якості міри нелінійності функції $\theta(f)$ зручно

використовувати максимальну різницю $\Delta\tau_{\max} = \tau(f_{8000}) - \tau(f_{125})$ значень групового часу затримки $\tau(f)$, обчислених для крайніх фільтрів гребінки, тобто фільтрів з центральними частотами $f_0 = 125$ Гц та $f_0 = 8000$ Гц.

При об'єктивному оцінюванні якості мовних сигналів використано фрагменти тривалістю 1 хвилина кожний мовних сигналів для 4 дикторів-жінок і 4 дикторів-чоловіків, які читають текст російською мовою з юридичної тематики. Запис сигналів проведено на кафедрі акустики Національного технічного університету «Київського політехнічного інституту імені Ігоря Сікорського», в заглушеному приміщенні з часом реверберації 0,15 с.

При об'єктивному оцінюванні якості музичних сигналів використано фрагменти восьми музичних творів тривалістю 30-45 секунд кожен. При цьому половина творів належала до жанру класичної музики («Ave Maria» Дж. Каччіні, «Етюд №4, ор. 10, Ф. Шопена», 5-а симфонія П. Чайковського, увертюра «Фауст» Р. Вагнера), а половина – до жанру популярної музики («Mamma mia» ABBA, «She's Leaving Home» The Beatles, «Я піду в Далекі гори» К. Цісик, «Mademoiselle Hyde» L. Fabian).

Як мовні, так і музичні сигнали записано з частотою дискретизації 22050 Гц і бітовою глибиною 16 біт.

У суб'єктивному оцінюванні брало участь 32 людини віком 21-23 роки без вад слуху. Оцінювання якості сигналів вироблялося за шкалою DMOS, з використанням спеціально розробленої комп'ютерної програми, за допомогою якої слухачеві у випадковому порядку пред'являлися спотворені сигнали для $|\Delta\tau_{\max}| \approx 30, 50, 70$ та 90 мс, а також неспотворений сигнал ($\Delta\tau_{\max} = 0$).

На рис. 2.13 представлені усереднені, за слухачами, а також за окремими видами мови і музики, результати суб'єктивного оцінювання якості сприйняття мовита музики. Варто наголосити, що такі оцінки отримано вперше.

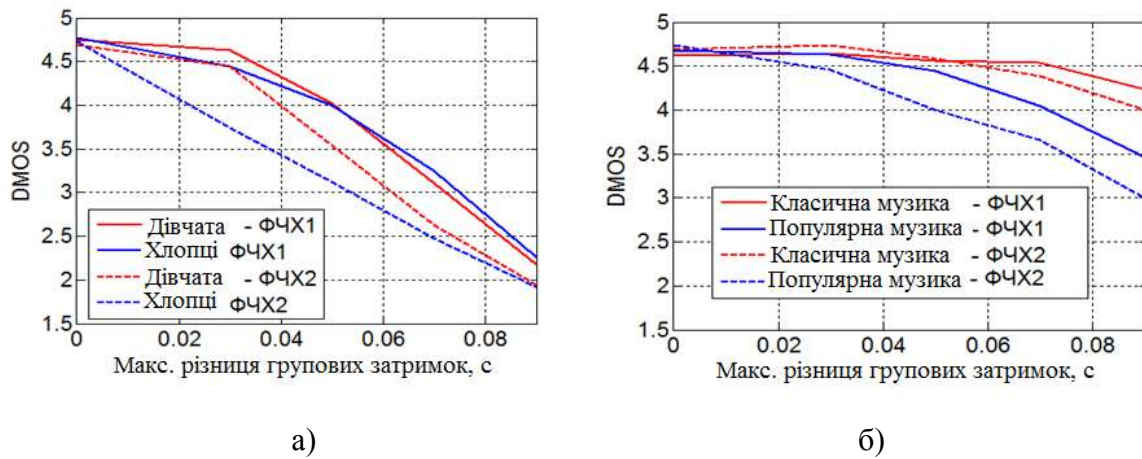


Рис. 2.13. Графіки суб'єктивних оцінок якості мови (а) і музики (б)

Перш за все, зазначимо, що фазові спотворення музичних сигналів значно менш помітні на слух, ніж для мовних сигналів. Дійсно, при $\Delta\tau_{\max} \approx 90$ мс оцінки якості мовних сигналів близькі до 2 балів за шкалою DMOS (рис. 2.13, а), тоді як оцінки якості музичних сигналів залишаються порівняно високими та вкладаються в інтервалі 3-4 бали (рис. 2.13, б).

Для практичних застосувань зручніше зафіксувати необхідну якість звукового сигналу на певному рівні та визначити порогові значення $\Delta\tau_{\max, DMOS \geq level}$, перевищення яких є неприпустимим. Наприклад, для рівня 4 бали шкали DMOS ($level = 4$) із графіків рис. 2.13 можна отримати:

$$\Delta\tau_{\max, DMOS \geq 4} = \begin{cases} 0.5 \text{ s} & \text{for speech, } \Delta\tau_{\max} < 0; \\ 0.1 \text{ s} & \text{for speech, } \Delta\tau_{\max} > 0; \\ 0.9 \text{ s} & \text{for music, } \Delta\tau_{\max} < 0; \\ 0.7 \text{ s} & \text{for music, } \Delta\tau_{\max} > 0. \end{cases}$$

Якщо усереднити всі криві на рис. 2.13, тоді можна отримати більш простий і легкий для запам'ятовування вираз:

$$\Delta\tau_{\max, DMOS \geq 4} = \begin{cases} 0.4 \text{ s} & \text{for speech;} \\ 0.8 \text{ s} & \text{for music.} \end{cases}$$

Зазначимо, що поведінка графіків рис. 2.13а узгоджується з попередніми висновками роботи [8] про те, що спотворення мовних сигналів для ФЧХ2 є помітнішим, ніж для ФЧХ1. Тому можна вважати підтвердженням висловлене в [8] припущення, що в ситуації «зростаючий час затримки» ($\Delta\tau_{\max} > 0$) при великих $\Delta\tau_{\max}$ відбувається «перетворення» відкритих складів на закриті. При цьому «співучість» мови знижується, аж до появи неприємного «дрижання». У ситуації «спадаючого часу затримки» ($\Delta\tau_{\max} < 0$) при досить великих значеннях $\Delta\tau_{\max}$ відбувається зворотне явище: закриті склади «перетворюються» у відкриті. У результаті загальна кількість відкритих складів зростає, і спотворення здаються меншими за рахунок збільшення «співучості» мови.

Для наочності зроблені висновки доповнимо моделлю у вигляді послідовності двох коротких імпульсів, по 40 мс кожний, розміщених у часі один за одним без паузи, з гармонічним заповненням різної частоти: частота заповнення першого імпульсу 125 Гц, а другого імпульсу – 8 кГц (рис. 2.14, а). Для мовних сигналів розглянута модель сигналу відповідає так званим закритим складам («ас», «ак» і т. п.). Пропускаючи такий сигнал через гребінку фільтрів з ФЧХ1, виявляємо, що на виході гребінки високочастотний та низькочастотний імпульси міняються місцями (рис. 2.14, б), тобто закриті немилозвучні склади перетворюються у відкриті милозвучні склади. Причина даного ефекту – істотна (70 мс) відмінність часів затримки для гармонік зазначених частот.

Аналогічну модель спотворень для гребінки фільтрів з ФЧХ2 наведено на рис. 2.15. Тут милозвучні відкриті склади перетворюються в немилозвучні закриті склади.

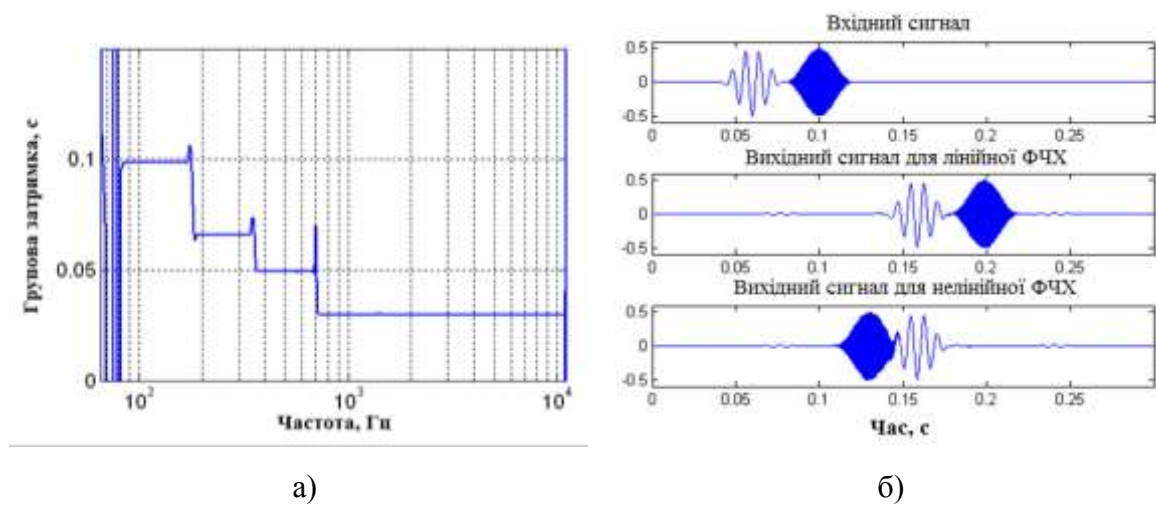


Рис. 2.14. Модель спотворень для ФЧХ1 при $\Delta\tau_{\max} \approx 70$ мс; групова затримка гребінки фільтрів (а), вхідні та вихідні сигнали (б).

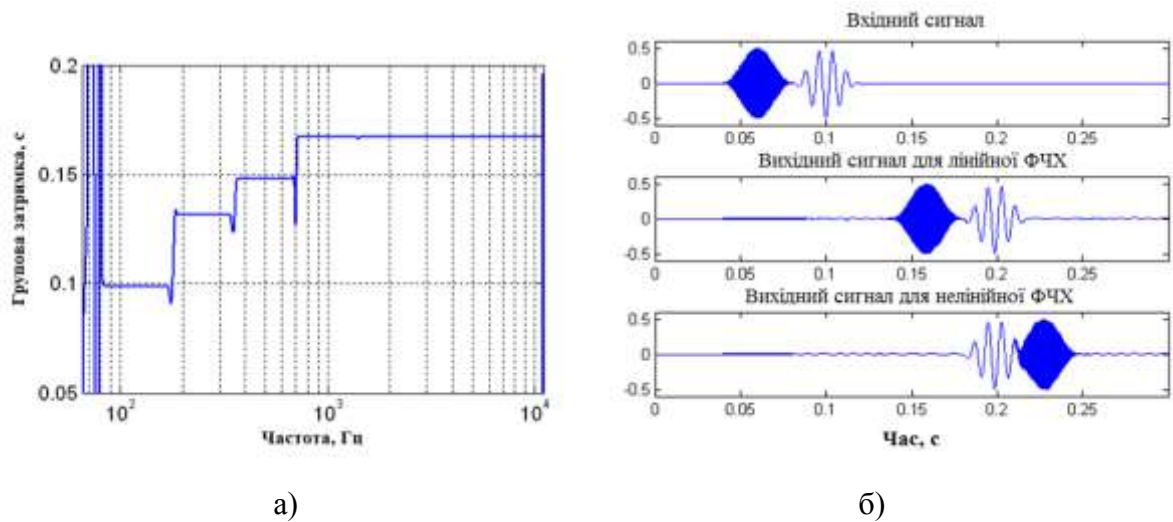


Рис. 2.15. Модель спотворень для ФЧХ2 при $\Delta\tau_{\max} \approx 70$ мс; групова затримка гребінки фільтрів (а), вхідні та вихідні сигнали (б).

Зрозуміло, наведене пояснення не може бути безпосередньо використано для зіставлення ФЧХ1 і ФЧХ2 при оцінюванні якості спотвореної музики (рис. 2.13, б). Проте, певна аналогія має місце, оскільки милозвучність мови зазвичай асоціюють з її вираженою мелодійністю.

Для об'єктивного оцінювання якості акустичних сигналів використано такі міри якості як SSNR, LSD, BSD та PESQ.

Результати об'єктивного оцінювання якості мовних сигналів представлені на рис. 2.16, а музичних сигналів – на рис. 2.17. На наведених графіках не відображено відмінність між кривими для ФЧХ1 і ФЧХ2. Причина в тому, що ці криві на графіках практично зливаються, оскільки відрізняються не більше ніж на 1-2%. Дуже цікавим та неочікуваним є те, що використані об'єктивні міри якості практично «не відчувають» відмінності форми спотворених сигналів, незважаючи на істотну відмінність фазових характеристик ФЧХ1 і ФЧХ2.

Несподіваність даного феномену фактично свідчить про недосконалість розглянутих в даній роботі об'єктивних мір якості.

Разом з тим, результати об'єктивного і суб'єктивного оцінювання досить добре узгоджуються в тому сенсі, що підтверджують меншу «чутливість» музичних сигналів до фазових спотворень.

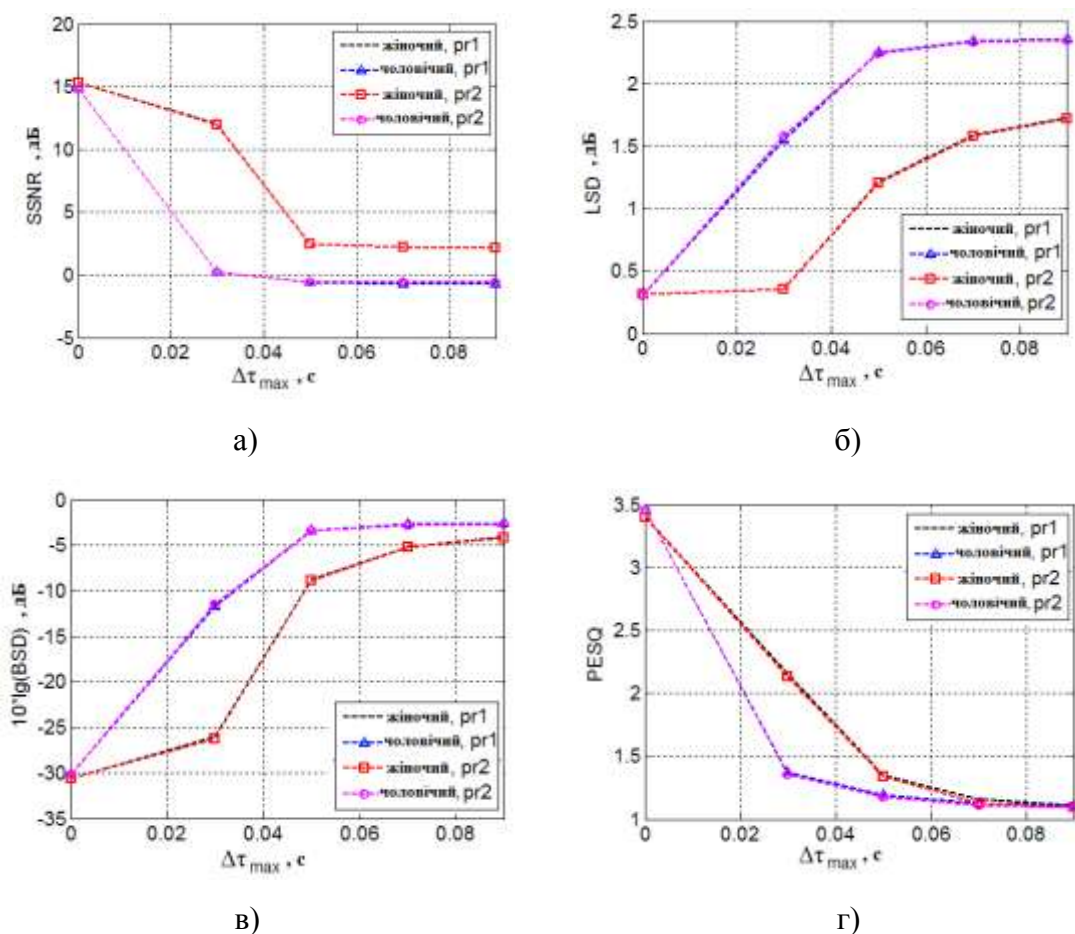


Рис. 2.16. Залежність усереднених за дикторами об'єктивних показників від $\Delta\tau_{\max}$ для мовних сигналів: SSNR(а), LSD(б), BSD(в), PESQ(г).

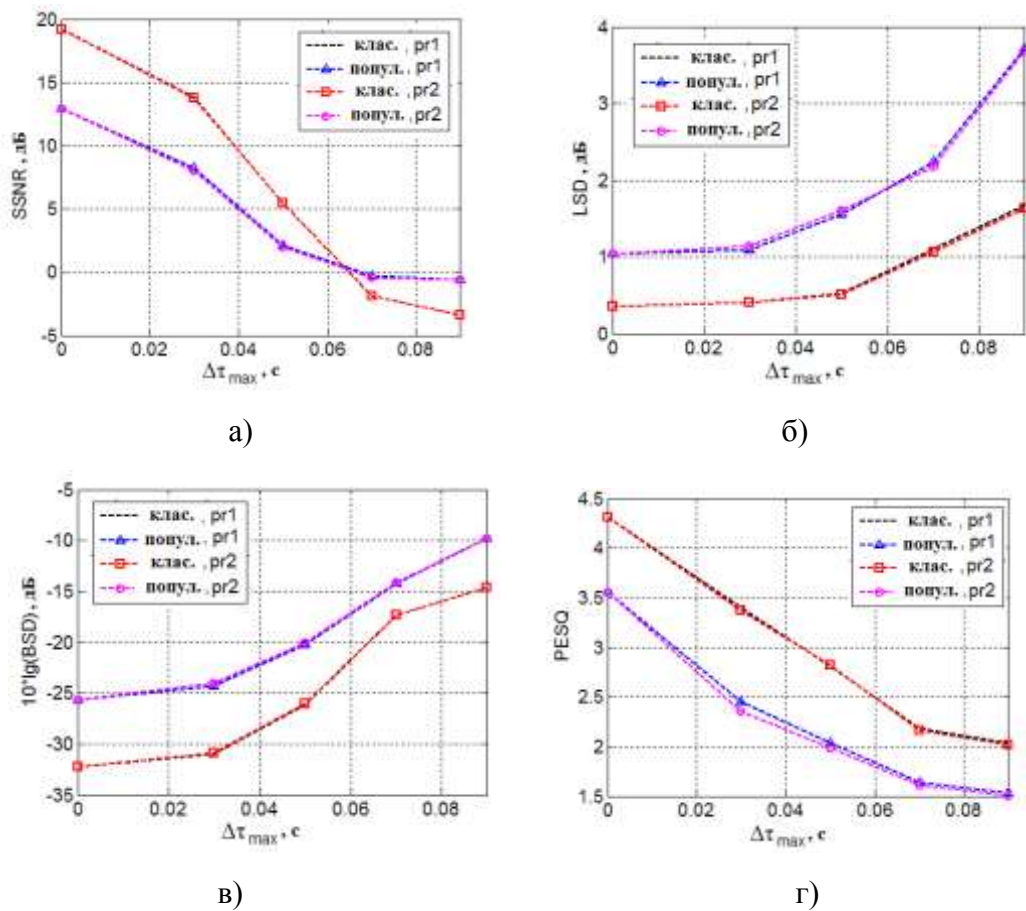


Рис. 2.17. Залежність усереднених за музичними творами об'єктивних показників від $\Delta\tau_{\max}$ для музичних сигналів: SSNR(а), LSD(б), BSD(в), PESQ(г).

Оскільки оцінки DMOS (рис. 2.13) та об'єктивні оцінки (рис. 2.16-2.17) є монотонними функціями максимальної різниці групових затримок $\Delta\tau_{\max}$, доцільно здійснити картування результатів DMOS та об'єктивного оцінювання. При цьому видаляється проміжна змінна $\Delta\tau_{\max}$, і дві криві об'єднуються в одну, як показано на рис. 2.18. Як бачимо, результуюча крива також є монотонною. Практична користь такого картування, як вже зазначалося раніше, полягає в можливості перерахунку результатів об'єктивного оцінювання в результати суб'єктивного оцінювання за шкалою DMOS.

Як можна бачити на рис. 2.18, деякі криві мають виражений нелінійний характер, що в певних ситуаціях може значно ускладнювати перерахунок результатів об'єктивного оцінювання в результати суб'єктивного оцінювання.

Такою є, наприклад, ситуація при $SSNR \approx 0$ для чоловічих голосів і при $SSNR \approx 2-2,5$ для жіночих голосів при оцінюванні якості мови (рис. 2.18, а). Аналогічною є ситуація при $SSNR \approx 0$ для популярної музики (рис. 2.18, б). З класичною музикою стан справи є трохи кращим, проте пояснюється це, головним чином, низькою чутливістю слухової системи людини до фазових спотворень класичної музики. Крім того, не слід нехтувати обмеженістю звукового матеріалу, що аналізувався.

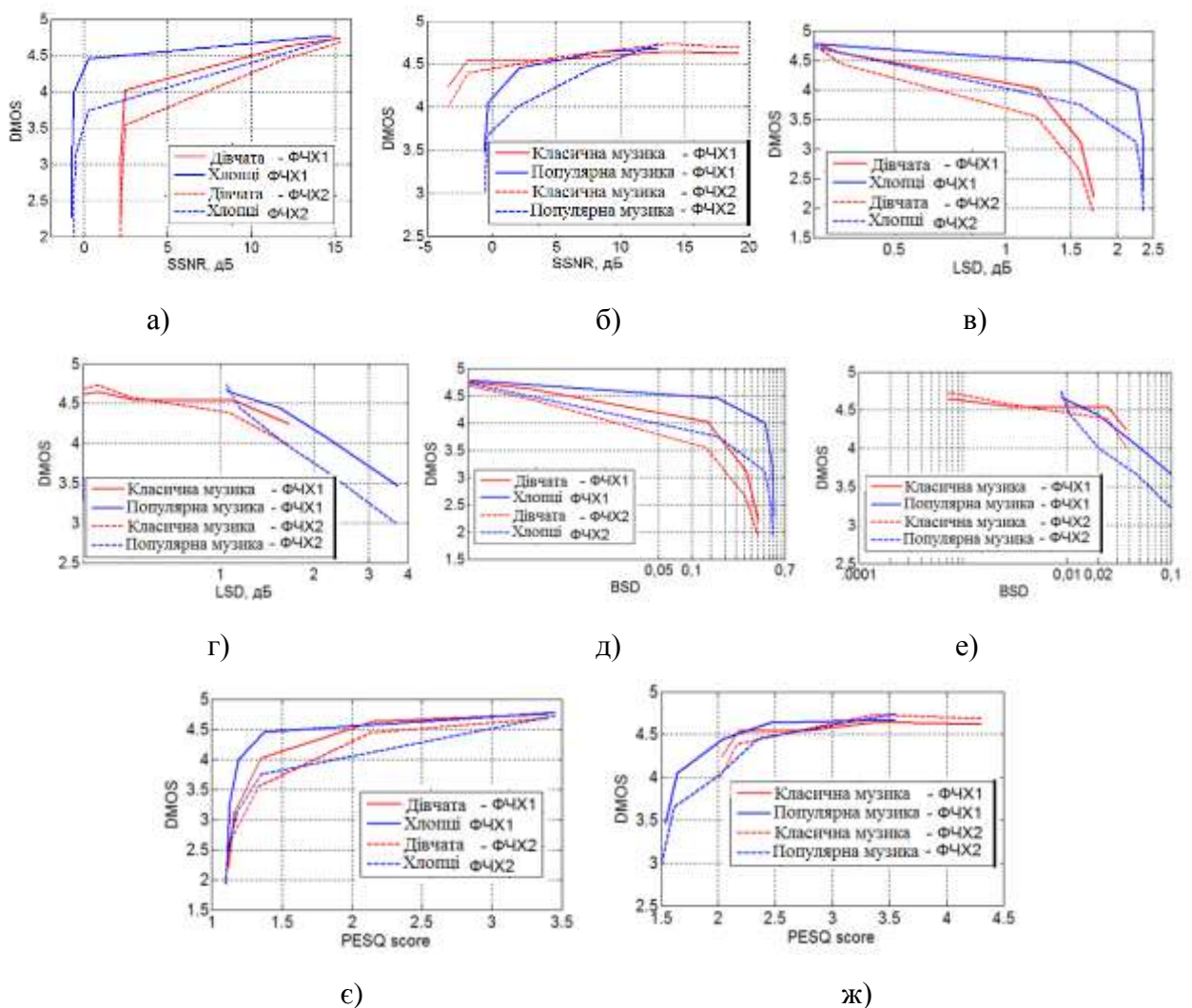


Рис. 2.18. Карти відповідності для оцінювання якості мови (а, в, д, є) та музики (б, г, е, ж)

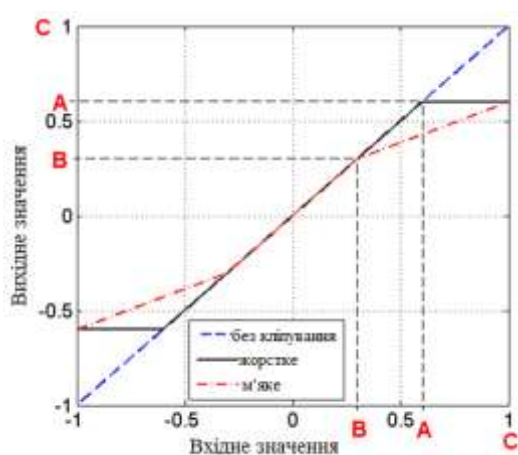
В силу монотонного характеру отриманих графіків, зазначена нелінійність карт відповідності не є перешкодою для їх практичного використання. Хоча,

зрозуміло, точність перерахунку об'єктивних оцінок в значення шкали DMOS є істотно зниженою на ділянках кривих з підвищеною крутизною.

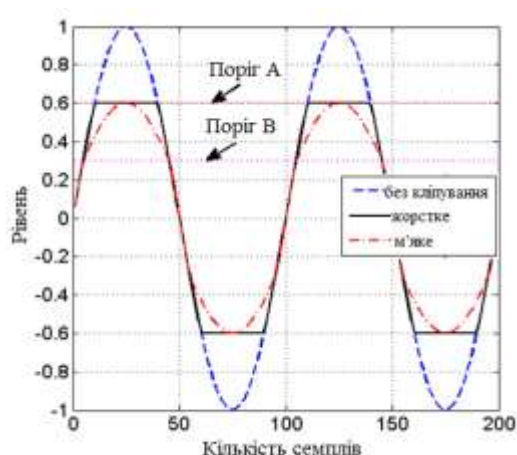
2.5. Оцінювання ступенів кліпування і якості кліпованої мови та музики.

У попередніх підрозділах розглянуто випадки лінійного спотворення сигналу системами передачі сигналу, такі як обмеження за смугою та фазові спотворення. У даному підрозділі розглянуто випадок найбільш розповсюдженого виду нелінійного спотворення сигналів, такого як кліпування.

Максимальне використання динамічного діапазону тракту передачі або запису мовних і музичних сигналів, з одного боку, є дуже бажаним, оскільки дозволяє мінімізувати дію шумової завади. Однак з іншого боку, такі дії пов'язані з небезпекою виникнення нелінійних спотворень, обумовлених кліпуванням, при якому великі миттєві значення сигналу замінюються деякою пороговою константою (жорстке кліпування) або близькими до неї значеннями (м'яке кліпування). Графічно механізм жорсткого та м'якого кліпування показано на рис. 2.19.



а)



б)

Рис. 2.19. Нелінійні передатні характеристики тракту (а) та форма кліпованого сигналу (б)

Для мінімізації втрат, викликаних кліпуванням сигналів, в тракти передачі й запису звукових сигналів вбудовують системи автоматичного регулювання підсилення (АРП) тракту, важливим елементом яких є підсистеми виявлення кліпування [11].

Мала ступінь кліпування супроводжується невеликими нелінійними спотвореннями сигналів, що не викликають негативної реакції з боку слухачів. Тому представляється розумним будувати алгоритм виявлення кліпування так, щоб рішення про наявність кліпування, сприйманого слухачами, передувало оцінюванню ступеня кліпування.

Ряд відомих способів виявлення кліпування базується саме на такому принципі, при цьому в більшості випадків в якості запобіжного ступеня кліпування пропонується використовувати ступінь відмінності форми або параметрів густини розподілу ймовірностей (PDF – probability density function) аналізованого і неспотвореного сигналів [12-19].

У даній дисертаційній роботі перевірено можливість використання ексцесу на роль міри ступеня кліпування стосовно мовних та музичних сигналів, а також продемонстровано перспективність ще двох мір, отриманих в результаті нескладних функціональних перетворень ексцесу.

2.5.1. Вступні зауваження

Підводячи підсумки наведеного у розділі 1 огляду відомих методів виявлення кліпування, відзначимо перспективність так званого «параметричного» напряму досліджень, згідно з яким рішення про ступінь кліпування приймається за результатами єдиного числового параметра, що характеризує форму PDF аналізованого сигналу. Представниками цього напряму є роботи [19], [21], що представляють евристичний алгоритм оцінювання однопараметричної міри ступеня кліпування, а також робота [20], де показана можливість використання ексцесу в ролі міри ступеня кліпування і міри якості кліпованих мовних сигналів. З огляду на попередній характер отриманих в [20]

результатів, доцільно уточнити їх, розглянувши додатково музичні сигнали. Крім того, оскільки можливі значення ексцесу належать інтервалу $[1, \infty)$, доцільно також розглянути конкурентні варіанти цієї міри у вигляді величин, отриманих шляхом нескладних функціональних перетворень ексцесу; можливі значення таких величин належать інтервалу $[0,1]$, що значно зручніше для інженерних застосувань.

2.5.2. Параметризація оцінок якості сигналів при жорсткому двосторонньому кліпуванні

У разі жорсткого двостороннього кліпування, спотворений сигнал $y(n)$ та вхідний сигнал $x(n)$ пов'язані співвідношенням

$$y(n) = \begin{cases} x(n), & |x| < A, \\ A \cdot \text{sign}[x(n)], & |x| \geq A, \end{cases} \quad (2.6)$$

де n – номер вибірки сигналу, A – границя кліпування ($0 < A < C = \max |x(n)|$), $\text{sign}(\cdot)$ – знакова функція, $|\cdot|$ – символ модуля.

Міру ступеня кліпування мовних сигналів було запропоновано в [10] і названо коефіцієнтом кліпування

$$R_{cl} = 2 \cdot \max(D_l, D_r) / D, \quad (2.7)$$

де D – різниця між максимальним і мінімальним значеннями сигналу, D_l та D_r – відстані між піками оцінки PDF. Рекомендації з оптимізації алгоритму оцінювання величини R_{cl} представлені в [11].

У [20] запропоновано міру ступеня кліпування і якості кліпованих мовних сигналів у вигляді нормованого центрального моменту четвертого порядку, іменовану для стислості в даній роботі, як і в роботах [14], [15], [16], ексцесом

$$\beta_4 = \frac{\mu_4}{(\mu_2)^2}, \quad (2.8)$$

де μ_k – центральний момент k -го порядку (очевидно, ексцес β_4 є тісно пов'язаним із коефіцієнтом ексцесу $\kappa = \beta_4 - 3$).

Принципову придатність мір (2.7) і (2.8) для виявлення кліпування не тільки мовних, а й музичних сигналів продемонструємо, використовуючи гістограмну оцінку некліпованого та кліпованого музичних сигналів, представлену на рис. 1 суцільними лініями. Штриховими лініями показані оцінки PDF некліпованого білого шуму з нормальним законом розподілу. Перед оцінюванням PDF, сигнали й шуми нормувалися по стандартному відхиленню, що дозволяє проводити коректне зіставлення форми отриманих оцінок PDF. Тривалість реалізацій сигналів становила близько 20 с, частота дискретизації 44100 Гц, при побудові оцінки PDF використано 100 бінів.

Порівняння наведених на рис. 2.20 значень оцінок β_4 та R_{cl} свідчить, що поява кліпування призводить до зменшення значень β_4 та до зростання значень R_{cl} .

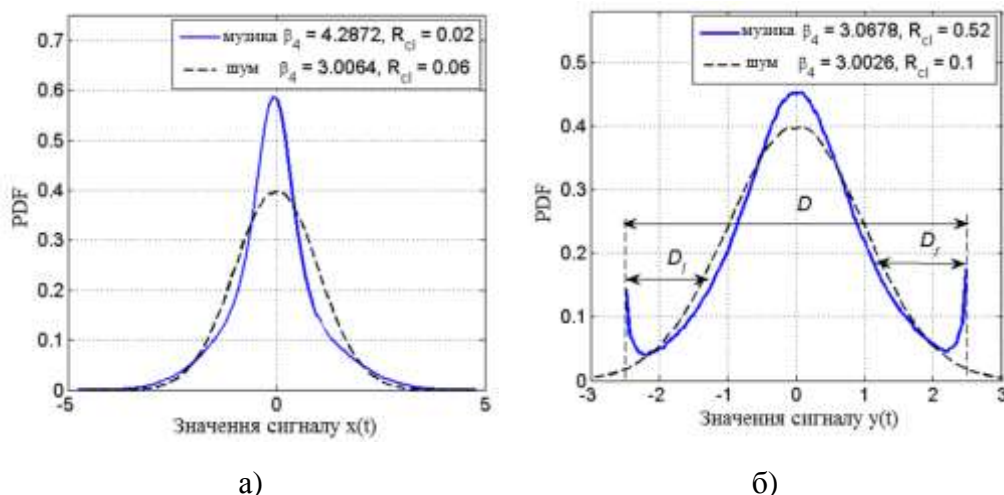


Рис. 2.20 Оцінки PDF, β_4 та R_{cl} для некліпованих (а) та кліпованих (б) музичних сигналів

Аналіз співвідношення (2.6) показує, що значення міри R_{cl} знаходяться в діапазоні $[0; 1]$, при цьому значення $R_{cl}=0$ відповідає ситуації відсутності кліпування.

Аналізуючи співвідношення (2.7), можна показати, що значення міри β_4 не можуть бути меншим за +1. Дійсно, позначаючи $Z = (X - E\{X\})/(V\{X\})^{1/2}$, де $E\{\cdot\}$ – символ математичного очікування, а $V\{\cdot\}$ – символ дисперсії, можна отримати [11]

$$\beta_4 = V\{Z^2\} + 1,$$

звідки $\beta_4 \in [1, \infty)$. Значення, що є близькими до $\beta_4=1$, відповідають сигналам, кліпованим в значній мірі. Хоча теоретично значення β_4 зверху необмежено, в разі реальних сигналів значення β_4 можуть досягати 30-50 для некліпованих музичних сигналів [12] и 7-12 для некліпованих мовних сигналів [13].

Оскільки зазначена «нечіткість» верхньої межі значень міри β_4 незручна в інженерних застосуваннях, замість β_4 логічно використовувати її функціональні перетворення:

$$\gamma_4 = 1/\beta_4, \quad \eta_4 = \sqrt{\gamma_4} = 1/\sqrt{\beta_4}, \quad (2.9)$$

діапазон можливих значень яких обмежений інтервалом $[0; 1]$, причому $\gamma_4=0$ відповідає ситуації відсутності кліпування, як і при використанні параметра R_{cl} .

Досліджуючи на реальних мовних і музичних сигналах можливості мір β_4 , $\gamma_4=1/\beta_4$ та $\eta_4=1/\sqrt{\beta_4}$, доцільно зіставити їх з такими для міри R_{cl} , тим більше що в [16] зафіксовано низьку придатність міри R_{cl} для оцінювання ступеня кліпування.

2.5.3. Експериментальні дослідження мір ступеня кліпування

Мета експериментальних досліджень полягала в з'ясуванні здатності величин β_4 , $\gamma_4 = 1/\beta_4$, $\eta_4 = 1/\sqrt{\beta_4}$ та R_{cl} грати роль об'єктивних мір ступеня кліпування і мір якості мовних та музичних сигналів. Для експериментальних досліджень використовувалися записи сигналів тривалістю 15-20 с.

Запис мовних сигналів виконували в заглушеному приміщенні з часом реверберації 0,15 с при відношенні сигнал-шум 38 дБ. Один і той самий текст з юридичної тематики зачитувався 8 дикторами (4 чоловіки і 4 жінки) в нормальному темпі, але з мінімальними мовними паузами. Всі мовні сигнали були записані з частотою дискретизації 22050 Гц і бітовою глибиною 16 біт.

Музичні сигнали представляли собою фрагменти 8 музичних творів, половина яких належала жанру «класична музика» («Ave Maria» Дж. Каччіні, «Етюд №4, ор. 10, Ф. Шопена», 5-а симфонія П. Чайковського, увертюра «Фауст» Р. Вагнера), а друга половина – жанру «популярна музика» («Mamma mia» ABBA, «She's Leaving Home» The Beatles, «Я піду в Далекі гори» К. Цісик, «Mademoiselle Hyde» L. Fabian). Всі сигнали записані з частотою дискретизації 44100 Гц та бітовою глибиною 16 біт.

При моделюванні жорстко кліпованих сигналів (2.5), управління ступенем кліпування здійснювалося за допомогою невід'ємної величини [16]

$$k = 20 \lg(\max |x(n)|/A), \quad (2.10)$$

причому значення $k = 0$ відповідає ситуації відсутності кліпування.

Експериментальні дослідження проводилися в наступному порядку:

- 1) акустичний сигнал піддавався жорсткому кліпуванню відповідно до (2.5), при цьому значення ступеня кліпування (6) змінювалися в діапазоні $k = 0 \dots 20$ дБ;
- 2) здійснювалося суб'єктивне оцінювання якості сигналів, різного ступеня кліпування;

- 3) здійснювалося оцінювання величин β_4 , $\gamma_4 = 1/\beta_4$, $\eta_4 = 1/\sqrt{\beta_4}$ та R_{cl} для сигналів, різного ступеня кліпування;
- 4) будувалися карти відповідності результатів суб'єктивного і об'єктивного аналізу сигналів різного ступеня кліпування.

Суб'єктивне оцінювання якості кліпованих сигналів здійснювалося з використанням 5-бальної шкали DMOS шляхом порівняння на слух якості спотвореного і чистого сигналів, причому 5 балів ставилося тоді, якщо на слух спотворення не відчувалися, а 1 бал ставилося тоді, коли звучання спотвореного сигналу дуже дратувало слухачів. Якість мовних сигналів оцінювалося 6 слухачами, а в оцінюванні якості музичних сигналів брало участь 9 слухачів, ніхто зі слухачів не мав вад слуху.

При оцінюванні величин β_4 , $\gamma_4 = 1/\beta_4$ и $\eta_4 = 1/\sqrt{\beta_4}$ використовувалася незміщена оцінка величини β_4 :

$$\overline{\beta_4} = \frac{N-1}{(N-2)(N-3)} \times \left((N+1) \cdot \overline{\beta_4'} - 3(N-1) \right) + 3,$$

$$\overline{\beta_4'} = \frac{\frac{1}{N} \sum_{n=1}^N (y(n) - \overline{m_y})^4}{\left(\frac{1}{N} \sum_{n=1}^N (y(n) - \overline{m_y})^2 \right)^2}, \quad \overline{m_y} = \frac{1}{N} \sum_{n=1}^N y(n),$$

де N – кількість вибірок аналізованого сигналу, m – символ математичного очікування, а риса над символами означає оцінку.

Оцінювання величини R_{cl} здійснювалося за попередньо побудованою гістограмою (100 бінов), відповідно до наведеного в [16] алгоритму.

Вперше отримані результати суб'єктивного оцінювання якості кліпованої мови і музики від ступеня кліпування k представлено на рис. 2.21 у вигляді залежностей $DMOS(k)$. Як бачимо, якість кліпованих сигналів залишається суб'єктивно високою ($DMOS \geq 4,5$) при умові $k \leq 5$ дБ для мови та $k \leq 7$ дБ для

музики. Суб'єктивно гарна якість мовних сигналів ($4 \leq DMOS < 4,5$) досягається при $5 < k \leq 8$ дБ для мови та $7 < k \leq 10$ дБ для музики. Узагальнюючи отримані результати, можна попередньо зробити висновок, що слухова система людини допускає ступінь кліпування музичних сигналів на 2-3 дБ більшу, ніж для мовних сигналів. Як бачимо, аналогічна ситуація спостерігалася й при лінійних спотвореннях акустичних сигналів.

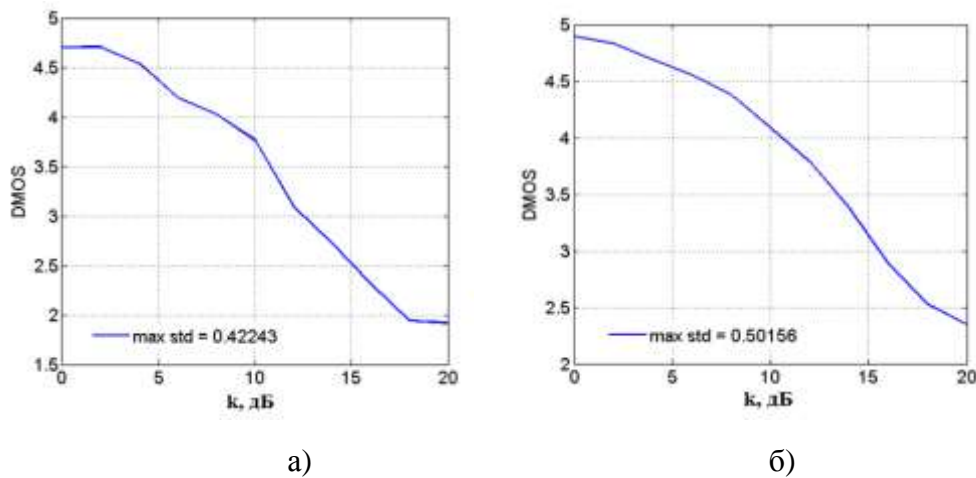
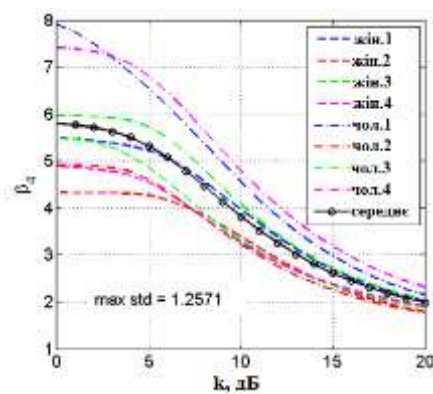


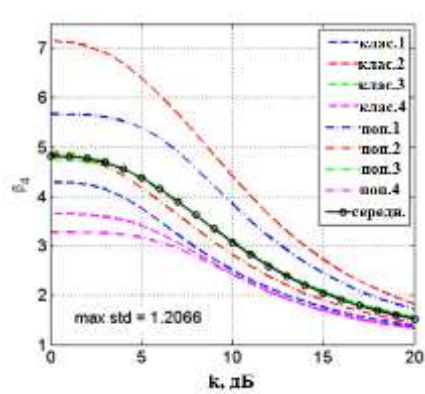
Рис. 2.21. Залежності $DMOS(k)$ для мови (а) та музики (б)

Результати оцінювання величин β_4 , $\gamma_4 = 1/\beta_4$, $\eta_4 = 1/\sqrt{\beta_4}$ та R_{cl} у вигляді залежностей $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$, $\overline{\eta_4}(k)$ та $\overline{R_{cl}}(k)$ представлені на рис. 2.22-2.25, де $\max std$ – максимальне значення стандартного відхилення отриманих оцінок.

Залежності $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$ та $\overline{\eta_4}(k)$ змінюються дуже мало на інтервалі $0 \leq k \leq 5$ дБ, тобто при низькому ступені кліпування, при якому суб'єктивно якість мови і музики залишається високою. Тим часом, на цікавому для практичного використання інтервалі $5 < k \leq 15$ дБ, де суб'єктивно якість мови знижується з 4,5 балів до 2 балів за шкалою DMOS, залежно $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$ та $\overline{\eta_4}(k)$ змінюються в помірному темпі і з практично постійною швидкістю. Ця сталість і помірність швидкості зміни означає, що величини β_4 , $\gamma_4 = 1/\beta_4$ та $\eta_4 = 1/\sqrt{\beta_4}$ є хорошими претендентами на роль мір ступеня кліпування.

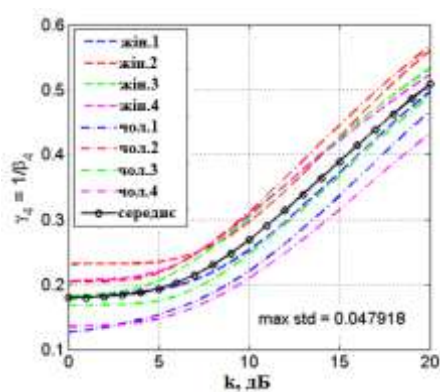


а)

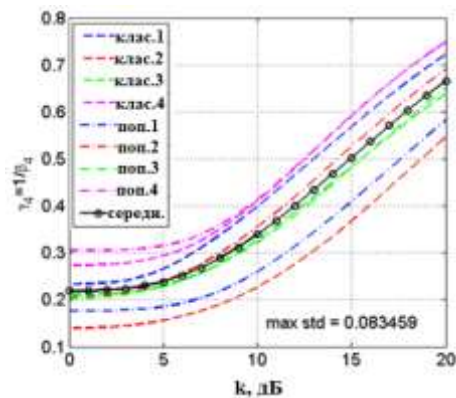


б)

Рис. 2.22. Залежності $\bar{\beta}_4(k)$ для мови (а) та музики (б)

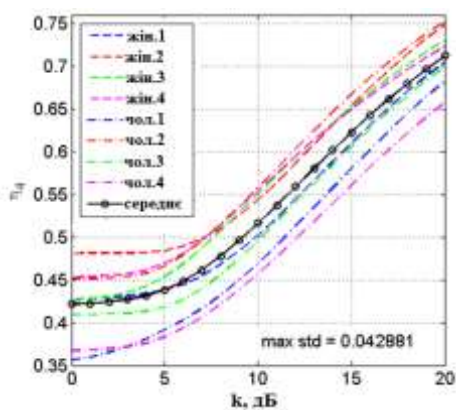


а)

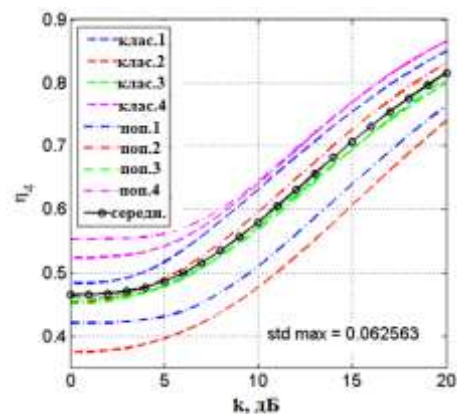


б)

Рис. 2.24. Залежності $\bar{\gamma}_4(k)$ для мови (а) та музики (б)



а)



б)

Рис. 2.25. Залежності $\bar{\eta}_4(k)$ для мови (а) та музики (б)

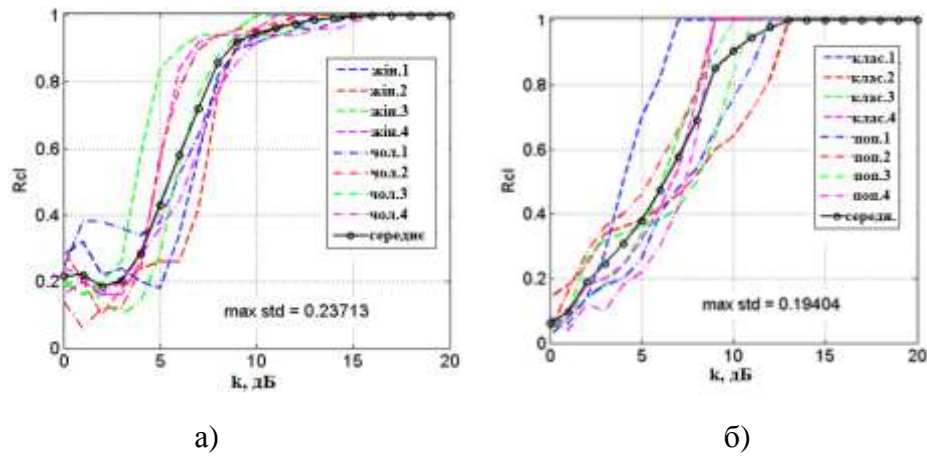


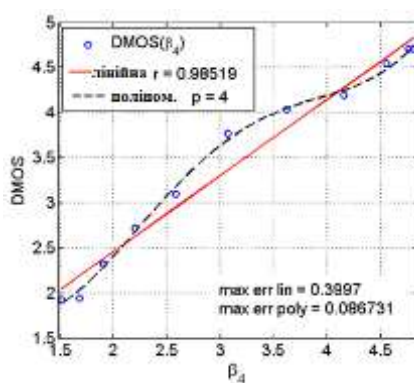
Рис. 2.26. Залежності $\overline{R_{cl}}(k)$ для мови (а) та музики (б)

Залежність $\overline{R_{cl}}(k)$, змінюється дуже швидко в інтервалі $0 \leq k \leq 8$ дБ, в межах якого суб'єктивно якість мови і музики залишається хорошою. Однак на інтервалі $8 < k \leq 15$ дБ, де якість сигналів змінюється від хорошої (4 бали за шкалою DMOS) до поганої (2 бали), залежність $\overline{R_{cl}}(k)$ змінюється незначно. Ці результати цілком узгоджуються з висновком роботи [16] про практичну непридатність величини R_{cl} для оцінювання ступеня кліпування мовних сигналів в умовах сильного кліпування. Більш того, як бачимо, цей висновок справедливий і для музичних сигналів.

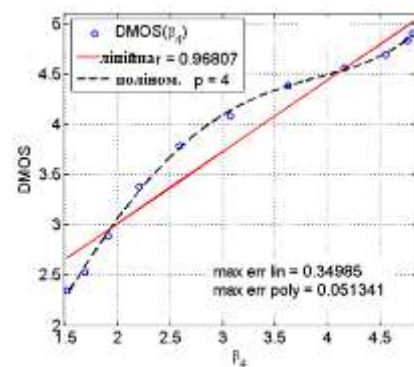
2.5.4. Міри ступеня кліпування як міри якості кліпованих сигналів

Якщо в інженерному застосуванні ставиться завдання об'єктивного оцінювання якості кліпованих мовних і музичних сигналів, тоді доцільно використовувати карти відповідності $DMOS(\beta_4)$, $DMOS(\gamma_4)$, $DMOS(\eta_4)$ та $DMOS(R_{cl})$, які наведені на рис. 7-10 і побудовані шляхом зіставлення значень залежностей $DMOS(k)$ (рис. 2.21) та усереднених значень залежностей $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$, $\overline{\eta_4}(k)$ та $\overline{R_{cl}}(k)$ (рис. 3-6) для однакових значень параметра k . Практичне значення карт відповідності полягає в можливості здійснити калібрування шкал значень величин β_4 , γ_4 та η_4 .

На рис. 2.27-2.30 кружечками показані 11 значень карт відповідності, отриманих для k , що змінюється від 0 дБ до 20 дБ з кроком 2 дБ. Суцільними лініями показані результати лінійної апроксимації цих значень, а пунктирними лініями – результати поліноміальної апроксимації. У верхній частині кожного графіка вказані оцінки коефіцієнта кореляції r , а також порядки p поліномів для випадку поліноміальної апроксимації. Значення максимальних помилок лінійної і поліноміальної апроксимації наведені в нижній частині кожного графіка (max err lin та max err poly, відповідно).

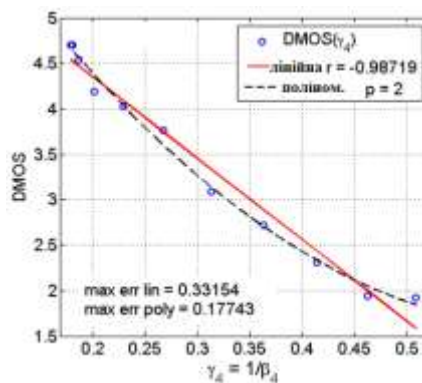


а)

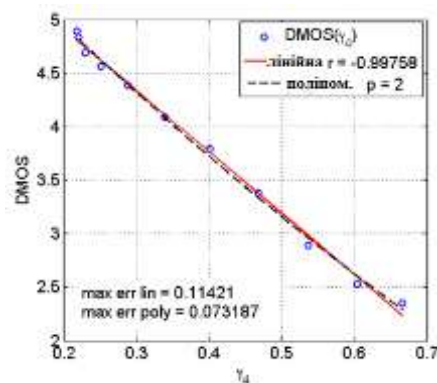


б)

Рис. 2.27. Карті відповідності $DMOS(\beta_4)$ для мови (а) та музики (б)

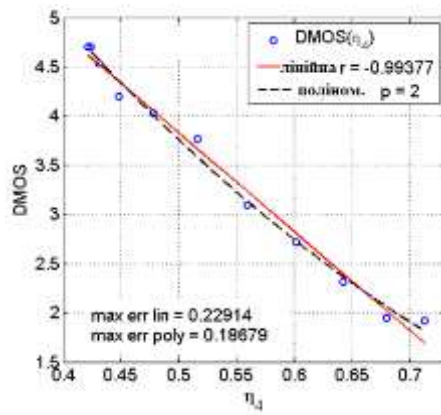


а)

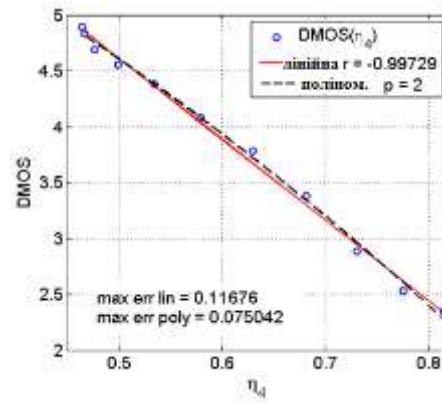


б)

Рис. 2.28. Карті відповідності $DMOS(\gamma_4)$ для мови (а) і музики (б)

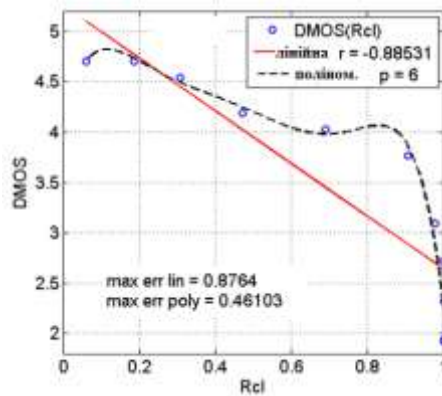


а)

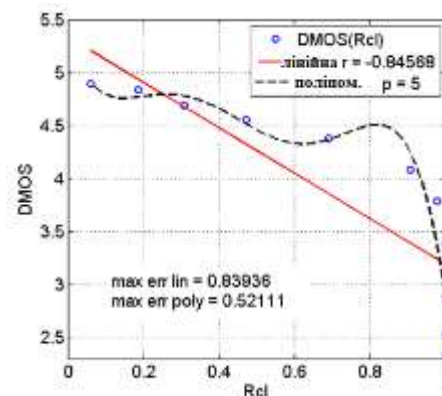


б)

Рис. 2.29. Карті відповідності $DMOS(\eta_4)$ для мови (а) та музики (б)



а)



б)

Рис. 2.30. Карті відповідності $DMOS(R_{cl})$ для мови (а) та музики (б)

Аналіз наведених карт відповідності дозволяє зробити висновок, що в силу високої корельованості (табл. 2.3) результатів об'єктивного та суб'єктивного оцінювання якості кліпованих сигналів, при поліноміальній апроксимації залежностей $DMOS(\eta_4)$ достатньо використовувати поліноми першого порядку (підкреслимо, що порядок полінома позначений символом p). Для забезпечення малої помилки апроксимації $DMOS(\gamma_4)$ можна використовувати поліноми першого (музика) або другого (мова) порядку. Аналітичний опис залежностей $DMOS(\beta_4)$ є задовільним при використанні поліномів більш високого, 4-го порядку. Що стосується залежності $DMOS(R_{cl})$, мала помилка апроксимації може бути досягнута лише при використанні поліномів 5-6-го порядків, проте навіть у

цьому випадку якість апроксимації важко визнати задовільною через немонотонність отриманої карти відповідності.

Значення коефіцієнтів апроксимуючих поліномів:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots$$

для залежностей $DMOS(\beta_4)$, $DMOS(\gamma_4)$, $DMOS(\eta_4)$ та $DMOS(R_{cl})$ у випадку мовних сигналів наведені в табл. 2.4, а в разі музичних сигналів – в табл. 2.5.

Таблиця 2.3. Коефіцієнти кореляції

Міри		β_4	γ_4	η_4	R_{cl}
	Мова	0,985	-0,987	-0,994	-0,885
	Музика	0,968	-0,998	-0,997	-0,846

Таблиця 2.4. Коефіцієнти апроксимуючих поліномів для мови

Коефіцієнти	β_4	γ_4	η_4	R_{cl}
	$p=4$	$p=2$	$p=1$	$p=5$
a_0	5,726	7,523	8,856	4,087
a_1	-7,883	-18,536	-10,062	16,621
a_2	5,126	14,481	-	-132,091
a_3	-1,203	-	-	460,869
a_4	0,098	-	-	-834,493
a_5	-	-	-	752,484
a_6	-	-	-	-265,219

Таблиця 2.5. Коефіцієнти апроксимуючих поліномів для музики

Коефіцієнти	β_4 $p=4$	γ_4 $p=1$	η_4 $p=1$	R_{cl} $p=5$
a_0	-0,069	6.053	8,210	5,360
a_1	0,909	-5.730	-7,204	-11,366
a_2	0,866	-	-	76,019
a_3	-0,337	-	-	-217,702
a_4	0,034	-	-	265,063
a_5	-	-	-	-114,515

Похибку апроксимації залежності $DMOS(R_{cl})$ можна істотно зменшити, якщо використовувати кусочно-лінійну апроксимацію (рис. 2.31).

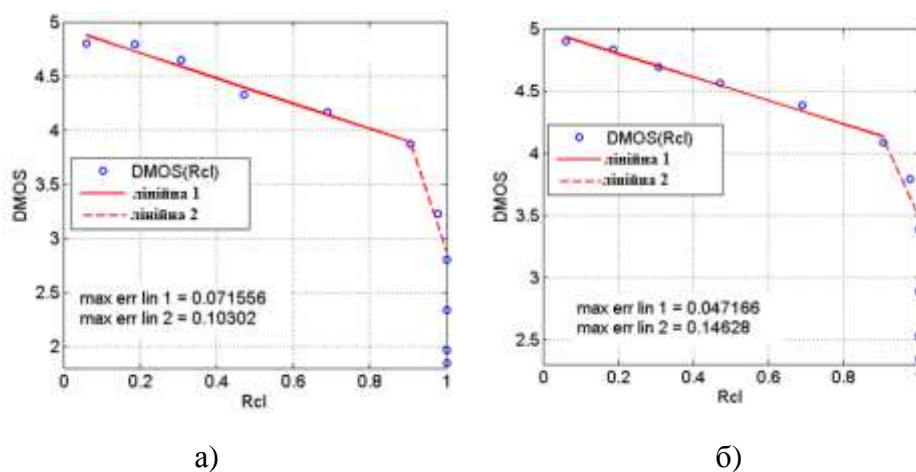


Рис. 2.31. Кусочно-лінійна апроксимація залежностей $DMOS(R_{cl})$ для мови (а) и музики (б)

Для мовних сигналів рівняння кусочно-лінійних залежностей, представлених на рис. 2.31, мають вигляд:

$$DMOS = \begin{cases} 4.95 - 1.17 \cdot Rcl, & 0 \leq Rcl \leq 0.9; \\ 13.6 - 10.72 \cdot Rcl, & 0.9 < Rcl \leq 1. \end{cases} \quad (2.11)$$

Для музичних сигналів аналогічні рівняння мають вигляд:

$$DMOS = \begin{cases} 4.99 - 0.95 \cdot Rcl, & 0 \leq Rcl \leq 0.9; \\ 10.11 - 6.62 \cdot Rcl, & 0.9 < Rcl \leq 1. \end{cases} \quad (2.12)$$

Співвідношення (2.10) – (2.12) і графіки рис. 2.31 добре узгоджуються з висновком [19] про низьку придатності міри R_{cl} для оцінювання ступеня кліпування, що дозволяє поширити цей висновок і на завдання оцінювання якості кліпованих сигналів. Навпаки, характер залежностей $DMOS(\beta_4)$, $DMOS(\gamma_4)$ та $DMOS(\eta_4)$ свідчить про принципову придатність ексцесу β_4 та його функціональних перетворень $\gamma_4 = 1/\beta_4$, $\eta_4 = 1/\sqrt{\beta_4}$ для використання не тільки в ролі мір ступеня клішування, але й в ролі мір якості кліпованих звукових сигналів.

Зазначимо, що в даній роботі результати вперше отримані на основі усереднених залежностей $DMOS(k)$, $\overline{\beta_4}(k)$, $\overline{\gamma_4}(k)$, $\overline{\eta_4}(k)$, $DMOS(\beta_4)$, $DMOS(\gamma_4)$ та $DMOS(\eta_4)$. В перспективі доцільно дослідити розподіли граничних значень мір клішування, хоча зробити це вкрай важко через великий об'єм експериментальних досліджень.

Можна вважати експериментально доведеним, що ексцес, зворотну йому величину, а також корінь квадратний з величини, зворотної до ексцесу, можна використовувати як об'єктивні міри ступеня кліпування і міри якості кліпованих мовних та музичних сигналів. Істотною перевагою запропонованих мір є можливість їх застосування за відсутності інформації про неспотворені сигнали.

Також значною перевагою є відсутність необхідності у попередньому оцінюванні щільності ймовірностей аналізованих сигналів.

Вперше отримані карти відповідності між запропонованими об'єктивними оцінками та суб'єктивними оцінками якості сигналів є гладкими монотонними функціями, що спрощує їх аналітичний опис і подальше використання в інженерних застосуваннях. Використання отриманих карт відповідності дозволяє обґрунтувати попередні висловлювання щодо низької ефективності конкурентної міри у вигляді коефіцієнта кліпування, в порівнянні з запропонованими мірами, при високому ступені кліпування мовних і музичних сигналів.

Нарешті, результати суб'єктивного оцінювання якості кліпованих сигналів свідчать, що слухова система людини допускає ступінь кліпування музичних сигналів на 2-3 дБ більшу, ніж для мовних сигналів, що слід враховувати при розробці інженерних додатків.

Висновки до розділу 2

1. Виконано уточнюючий аналіз особливостей об'єктивних мір SSNR і LSD якості мовних і музичних сигналів на прикладі сигналів, обмежених за смугою частот, що дозволило дати пояснення виявленим особливостям вказаних мір і виробити ряд рекомендацій, спрямованих на їх практичне використання.
2. Отримано нові залежності об'єктивних та суб'єктивних оцінок якості мовних та музичних сигналів, обмежених за смугою частот, від ширини цієї смуги частот. Зокрема, показано, що інтерполяція сигналів, що тестуються, дозволяє підвищити достовірність оцінювання з використанням показника SSNR, а саме: для мовних сигналів частоту дискретизації інтерпольованих сигналів достатньо збільшити вдвічі, тоді як для музичних сигналів необхідне збільшення у 5 разів. Крім того, показано, що можливими причинами порушення монотонності залежності

LSD від ширини смуги частот є нормування сигналів, що порівнюються, за дисперсією, висока чутливість LSD до особливостей спектру процесу, а також неідеальність прямокутної форми АЧХ використаних НЧ фільтрів.

3. Вперше побудовано карти відповідності між суб'єктивною мірою якості DMOS та об'єктивними мірами SSNR, LSD, BSD та PEAQ для ситуацій лінійного спотворення сигналів (обмеженість смуги частот та нелінійність фазової характеристики тракту) та нелінійного їх спотворення (кліпування), що робить можливим калібрування відповідних програмно-апаратних систем об'єктивного оцінювання якості музичних сигналів. Таке калібрування, на відміну від використання коефіцієнта кореляції, є більш точним, оскільки припускає існування нелінійної залежності між суб'єктивними та об'єктивними оцінками якості сигналів.
4. Уточнено попередні оцінки здатності слухової системи людини сприймати фазові спотворення мовних і музичних сигналів. Показано, що фазові спотворення мовних сигналів суб'єктивно і об'єктивно сприймаються сильніше, ніж для музичних сигналів.
5. Розглядаючи два види фазових спотворень сигналу: 1) групова затримка $\tau(f)$ є спадною функцією частоти; 2) групова затримка $\tau(f)$ є зростаючою функцією частоти, вперше показано, що слухова система людини є більш чутливою до другого виду фазових спотворень. Також вперше показано, що об'єктивні міри якості SSNR, LSD, BSD і PESQ не чутливі до різниці зазначених видів фазових спотворень.
6. Вперше отримано залежності суб'єктивних оцінок якості кліпованих мовних та музичних сигналів від ступеня їх кліпування, що створює необхідну базу для подальшого калібрування об'єктивних мір ступеня кліпування.
7. Вперше запропоновано об'єктивні міри ступеня кліпування мовних та музичних сигналів у вигляді ексцесу, оберненої до нього величини, а також кореня квадратного з величини, оберненої до ексцесу.

РОЗДІЛ 3. ОЦІНЮВАННЯ РОЗБІРЛИВОСТІ МОВИ В УМОВАХ ДІЇ ШУМУ І РЕВЕРБЕРАЦІЇ

У даному розділі порушується питання підвищення ефективності суб'єктивного оцінювання розбірливості української мови шляхом автоматизації такого оцінювання. Розроблено та протестовано прототипи таблиць звукосполучень, притаманних українській мові, а також розроблено комплекс комп'ютерних програм, призначених для полегшення процедури суб'єктивного оцінювання розбірливості звукосполучень. Дієвість створеного комплексу перевірено шляхом оцінювання розбірливості мови, спотвореної шумом та реверберацією. Крім того, виявлено умови, за яких можливо оцінювати розбірливість мови за спрощеними алгоритмами оцінювання якості мовних сигналів.

3.1. Автоматизація артикуляційних досліджень

Ключовим елементом розробленої в рамках даної дисертаційної роботи автоматизованої системи є спеціальний комплекс комп'ютерних програм та таблиці звукосполучень типу «приголосний-голосний-приголосний» для української мови. Низка розроблених комп'ютерних програм дозволяє змішувати записи звукосполучення з шумом та додавати реверберацію в різних співвідношеннях, програвати їх слухачам, фіксувати та обробляти результати прослуховування.

На рис. 3.1 наведено схему процедури артикуляційних випробувань, яка складається з п'яти етапів, при цьому лише третій етап прослуховування спотворених звукосполучень неможливо піддати автоматизації.

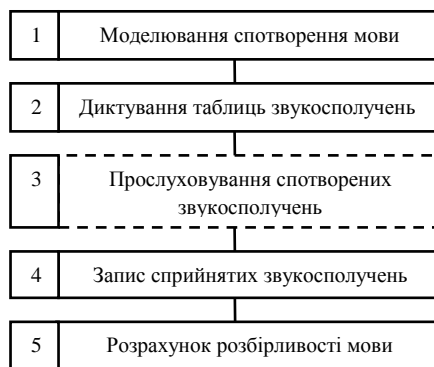


Рис. 3.1. П'ять етапів процедури артикуляційного випробування

Артикуляційні таблиці є важливим елементом артикуляційних тестів, від яких істотно залежить надійність результату. Фрагмент артикуляційних таблиць стандарту 50840-95 [24], призначених для неавтоматизованого варіанту артикуляційних тестів, показаний на рис. 3.2а. Аналогічний фрагмент таблиць для автоматизованої версії артикуляційних тестів із цього ж стандарту показаний на рис. 3.2б.

<div>заль фиц гет(д)ь ра(б)пъ</div> <div>ши(ы)н чтаф(в) буш(жъ) прёт(д)</div>	<div>__як/ __ак</div> <div>__юль/ __уль</div> <div>__юш/ __уш</div> <div>__ят/ __ат</div> <div>__вум</div>
а)	б)

Рис. 3.2. Фрагменти стандартних таблиць 50840-95, призначених для неавтоматизованих (а) та автоматизованих (б) варіантів артикуляційних тестів

На першому етапі процедури артикуляційних випробувань за допомогою спеціальної комп'ютерної програми моделюється спотворення звукосполучень, заздалегідь записаних в заглушеному приміщенні. При цьому враховується можливість спотворення звуку шумом та реверберацією. На другому етапі під керівництвом розробленої комп'ютерної програми виконується озвучення спотвореного звуку. На третьому етапі спотворений звук прослуховується, а на четвертому етапі слухач занотовує почуті звукосполучення за допомогою

клавіатури комп'ютера. На фінальному, п'ятому, етапі відбувається автоматичний підрахунок частки правильно сприйнятих звукосполучень.

Якщо введене з клавіатури в комп'ютер звукосполучення співпадає із еталонним звукосполученням, що міститься в одній із 3-х еталонних текстових таблиць, комп'ютерною програмою приймається рішення про правильність сприйняття слухачем озвученого складу. У кожній таблиці міститься 50 таких звукосполучень, і після закінчення прослуховування таблиць на дисплеї відображається значення частки правильно почутих звукосполучень.

Крім того, на дисплеї відображається набір еталонних та введених з клавіатури звукосполучень. Зазначимо, що таке відображення є опціональним в робочому режимі, але корисним з технологічної точки зору, оскільки полегшує налаштування еталонних таблиць та комп'ютерних програм в процесі їх розробки.

Хоча окрім таблиць односкладових слів в зазначеному стандарті є також таблиці багатоскладових слів та фраз, таблиці односкладових слів вважаються кращим тестовим матеріалом з огляду на важкість відгадування слів. Враховуючи цю особливість, за основу було взято таблиці односкладових стандартів 50840-95 [24] та було розроблено дев'ять прототипів артикуляційних таблиць [25], де кожна таблицю подано у кількох варіантах (рис. 3.3) для врахування можливості невизначеного сприйняття окремих фонем. Як видно із порівняння рис. 3.2 та 3.3, запропоновані артикуляційні таблиці було отримано шляхом адаптації до української мови російськомовних таблиць, призначених для неавтоматизованого варіанту артикуляційних тестів.

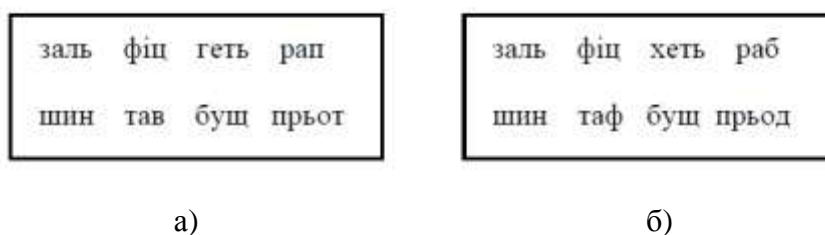


Рис. 3.3. Перші (а) і другі (б) фрагменти одного з розроблених артикуляційних тестів

3.2. Експериментальна перевірка дієздатності та ефективності автоматизованої системи артикуляційних досліджень

Моделювання сигналів, спотворених адитивним шумом та реверберацією виконувалося за загальним алгоритмом:

$$y(t) = x(t) \otimes h(t) + n(t) \quad (3.1)$$

де $x(t)$ – мовний сигнал, $h(t)$ – імпульсна характеристика приміщення, $n(t)$ – адитивний шум, \otimes – символ згортки. В окремому випадку дії виключно шумової завади алгоритм (3.1) можна спростити:

$$y(t) = x(t) + k \cdot n(t), \quad (3.2)$$

де $k = 10^{0.05(SNR-SNR_0)}$, $SNR = 10 \lg(D_x/D_n)$ є «початковим» значенням відношення сигнал-шум для чистої мови $x(t)$ та синтезованого «початкового» шуму $n(t)$, D_x та D_n є дисперсіями сигналу $x(t)$ та шуму $n(t)$, відповідно, SNR_0 є бажане SNR для суміші (3.2).

В окремому випадку сумісної дії шуму та реверберації, алгоритм (3.1) дещо ускладнюється з метою врахування особливостей структури імпульсної характеристики приміщення:

$$h(t) = h_e(t) + h_l(t) \quad (3.3)$$

$$h_e(t) = \begin{cases} h(t), & t \in 0..50 \text{ ms}; \\ 0, & t \notin 0..50 \text{ ms}, \end{cases}$$

$$h_l(t) = \begin{cases} h(t), & t > 50 \text{ ms}; \\ 0, & t \leq 50 \text{ ms}, \end{cases}$$

де $h_e(t)$ описує ранні відбиття, а $h_l(t)$ описує пізню реверберацію. Тоді «початкове» відношення сигнал-шум $SNR = 10 \lg D_{x_e}/D_n$, де D_{x_e} є дисперсією сигналу $x_e(t) = x(t) \otimes h_e(t)$.

3.2.1. Перший етап експериментальних досліджень

Для випробувань було розроблено дев'ять прототипів артикуляційних таблиць звукосполучень шляхом адаптації до української мови відповідних таблиць стандарту 50840-95 [24]. Кілька таких адаптованих таблиць наведено в Додатку 2 (табл. Д2.1).

При експериментальних випробуваннях цих таблиць та комплексу комп'ютерних програм використовувалися як чисті (неспотворені) мовленнєві сигнали, так і сигнали, спотворені шумом та реверберацією.

Записи чистих сигналів робилися в заглушеному приміщенні кафедри акустики та акустoeлектроніки НТУУ «КПІ ім. І. Сікорського». Для записів використовувалися мікрофон Superlux ECM 999, зовнішня звукова карта PRESONUS AudioBox USB та програма Audacity версії 2.2.0. Запис виконувався з частотою дискретизації 22050 Гц та бітовою глибиною 16 біт. Звукосполучення зачитувалися із застосуванням словесного оточення, наприклад, звукосполучення «няк» зачитувалося як «Запишіть няк тепер».

Прослуховування відбувалося для чотирьох ситуацій:

- 1) чиста мова;
- 2) мова, спотворена шумом;
- 3) мова, спотворена реверберацією;
- 4) мова, спотворена шумом та реверберацією.

У першому випадку кожним слухачем по чергово прослуховувалися записи звукосполучень з артикуляційних таблиць, кожна з котрих містила 50 звукосполучень.

У другому випадку прослуховувалися звукосполучення, спотворені адитивним шумом з відношеннями сигнал-шум -10, 0 та +10 дБ. При цьому використовувалися моделі білого, рожевого та коричневого шумів, маскувальна властивість яких вважається досить добре вивченою [28].

У третьому випадку моделювалася (шляхом згортки сигналів чистої мови з імпульсними характеристиками різних приміщень) реверберована мова для часів реверберації від 0,3 до 2,7 с, а в четвертому випадку розглядався випадок спільної дії шуму та реверберації.

Прослуховування виконувалося через навушники 26 слухачами (середній вік – 22 роки). Три слухачі із зазначеного загалу виконали роботу з відхиленнями від завдання, а саме, прослуховуючи сигнали через акустичні монітори (пару акустичних колонок). Тому одержані результати було вирішено розділити на 2 групи: прослуховування через навушники та прослуховування через акустичні монітори.

Результати обробки одержаних даних представлено на рис. 3.4-3.7, де наведено графіки залежності усереднених оцінок розбірливості мови, спотвореної шумом та реверберацією, для випадків прослуховування через навушники та через акустичні монітори.

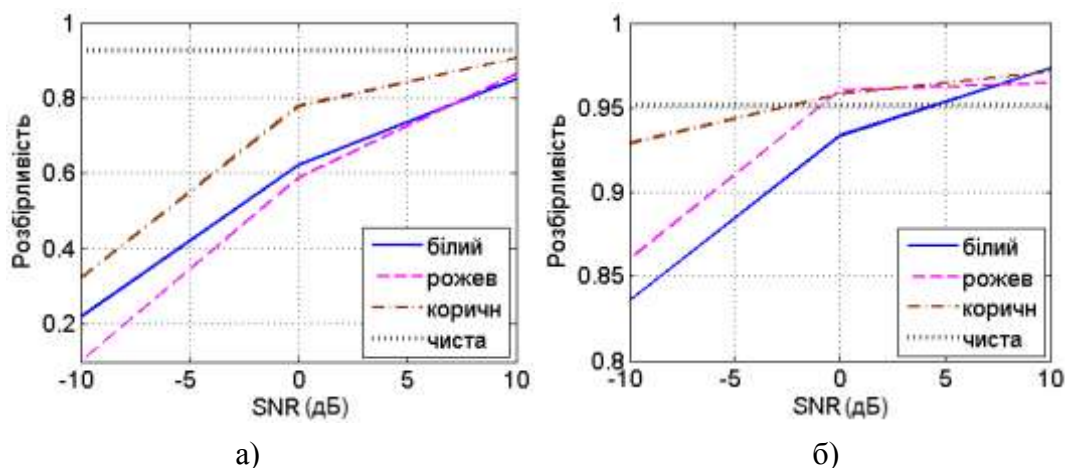


Рис. 3.4. Оцінки розбірливості зашумленої мови під час прослуховування через навушники (а) та монітори (б)

Аналіз графіків рис. 3.4 свідчить про те, що маскувальні властивості шуму залежать не тільки від відношення сигнал-шум та забарвленості шуму, але й від способу прослуховування. Як бачимо, при переході від прослуховування через навушники до прослуховування через акустичні монітори, розбірливість

піднялася з 10-30% до 85-93% при відношенні сигнал-шум мінус 10 дБ (стандартне відхилення оцінок розбірливості було близьким в середньому до 5%). Більш того, при відношенні сигнал-шум плюс 10 дБ розбірливість зашумленої мови навіть перевищила таку для чистої мови.

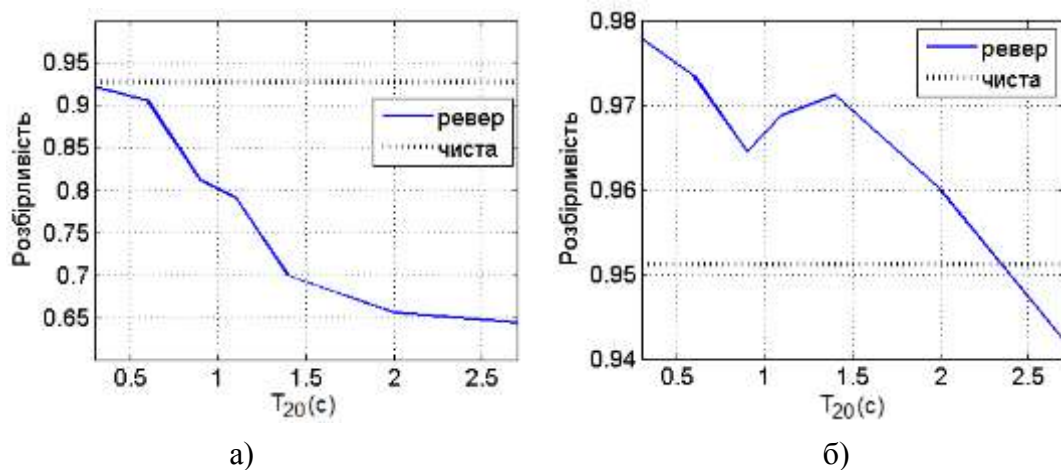


Рис. 3.5. Оцінки розбірливості мови з реверберацією під час прослуховування через навушники (а) та монітори (б)

Аналіз графіків рис. 3.5 підтверджує очікування щодо зниження розбірливості мови з ростом часу реверберації. Разом із тим, бачимо, що прослуховування через монітори значно підвищило розбірливість реверберованої мови. Так, розбірливість зросла з 65% до 94% для часу реверберації $T_{60}=2,7$ с, а для часу реверберації $T_{60}=0,3-2$ с розбірливість перевищила на 1-3% таку для чистої мови. Зазначимо, що в даному випадку стандартне відхилення оцінок розбірливості було досить великим і близьким до 10%, що пояснюється малою кількістю слухачів.

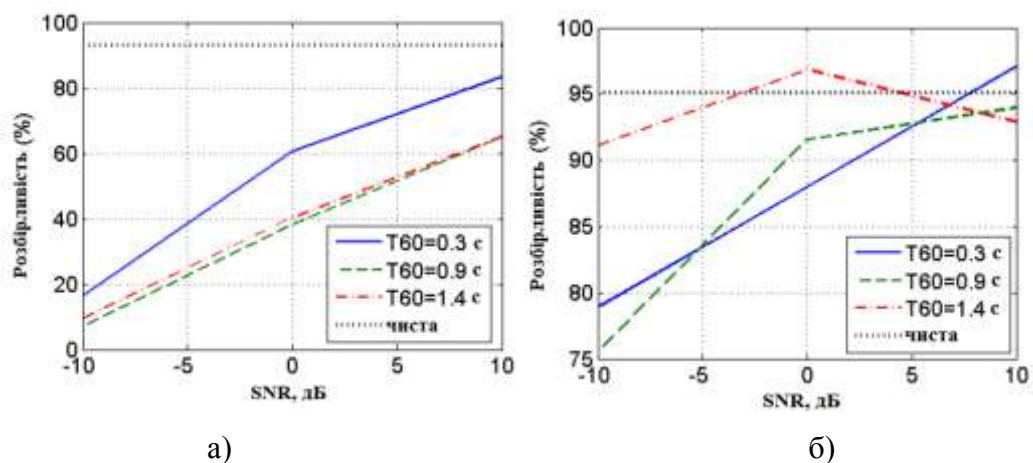


Рис. 3.6. Оцінки розбірливості мови, спотвореної шумом та реверберацією під час прослуховування через навушники (а) та монітори (б)

Аналіз результатів для випадку спільної дії шуму та реверберації (рис. 3.6) також свідчить про суттєве зростання розбірливості мови при прослуховуванні її через акустичні монітори (стандартне відхилення оцінок розбірливості не перевищувало 10%).

Коментуючи отримані результати, по-перше, відзначимо задовільну узгодженість результатів, представлених на рис. 3.4а, із такими в [27, 34, 35]. Однак, зазначимо, що маскувальні властивості білого шуму для малих (менше мінус 5 дБ) відношень сигнал-шум виявилися дещо гіршими, ніж в [34] та значно кращими, ніж в [35]. Можливою причиною таких порушень узгодженості може бути недостатня точність коефіцієнтів сприйняття, використаних в [34, 35]. Це припущення перевірено в пізніших експериментах. Випадки переважної дії реверберації (рис. 3.5) або спільної дії шуму та реверберації (рис. 3.6) на розбірливість мовлення досліджено менше [34], тому обмежимося зауваженням, що на якісному рівні зазначені графіки непогано узгоджуються з уявленнями про специфіку таких ситуацій.

По-друге, порівнюючи між собою результати прослуховування зашумлених сигналів на навушники та на акустичні системи (рис. 3.4, а і 3.4, б), бачимо, що при відношенні сигнал-шум мінус 10 дБ розбірливість сягає 80-90% при

використанні акустичних моніторів, тоді як при прослуховуванні на навушники така ж розбірливість може мати місце при відношеннях сигнал-шум не менше 5-7 дБ. Таким чином, доходимо дещо несподіваного висновку, що прослуховування зашумленого запису через акустичні монітори дозволило одержати виграш у відношенні сигнал-шум, що сягає 15-17 дБ.

Частково такий виграш можна пояснити дією ранніх відбиттів. Дійсно, в [34] знаходимо, що завдяки дії ранніх відбиттів виграш у відношенні сигнал-шум може сягати 6-9 дБ при прослуховуванні одного джерела звуку.

Спробуємо тепер пояснити одержаний нами «надлишок» виграшу, що становить приблизно 6-8 дБ. Якщо прослуховувати два гучномовця замість одного, як це має місце при прослуховуванні акустичних моніторів, відношення сигнал-шум на вході кожного вуха має збільшитися приблизно на 3 дБ, оскільки сигнали від випромінювачів складаються когерентно, а шуми від цих же випромінювачів складаються некогерентно. Додатковий виграш приблизно у 3 дБ можна пояснити бінауральним прослуховуванням [35]. Таким чином, доходимо висновку, що одержаний значний (15-17 дБ) виграш у відношенні сигнал-шум можна пояснити дією трьох чинників: ранні відлуння (9 дБ), подвоєння кількості джерел прямого звуку (3 дБ) та бінауральне прослуховування (3 дБ).

Як бачимо, після підсумовування окремих компонентів виграшу, одержуємо 15 дБ, тобто наближаємося «впритул» до значення фактичного виграшу. Видається правдоподібним припущення, що дія ранніх відбиттів в приміщеннях малого розміру (а саме такими були реальні приміщення, в яких розташовувалися акустичні монітори) є ще більш значною, ніж це вказано в роботі [28]. Для перевірки цього припущення експериментальні дослідження було в подальшому повторено – відповідні результати представлені в п. 3.2.2 даної дисертаційної роботи.

3.2.2. Другий етап експериментальних досліджень

Організація другої фази експерименту (п. 3.2.1) мала наступні відмінності:

Частота дискретизації становила 44100 Гц, аудіоредактор версії Audacity версії 2.1.3, на другій фазі було використано вдосконалене програмне забезпечення і набрано нову групу слухачів, яка складалася з 20 учнів у віці 20 років.

Корекція та модифікація програмного забезпечення було здійснено наступним чином:

- кількість текстових варіантів артикуляційних таблиць збільшено з двох до трьох;
- озвучення односкладових слів було виконано випадковим чином;
- стало можливим виправлення слухачами результатів введення з клавіатури;
- кількість візуальної інформації, що надається слухачеві на моніторі ПК після прослуховування артикуляційних таблиць, було мінімізовано;
- одні й ті ж артикуляційні таблиці кожен слухач прослуховував двічі: через навушники та через акустичні монітори.

Результати другого етапу експериментальних досліджень показані на рис. 3.7 і 3.8. У випадку шумового спотворення стандартне відхилення оцінок розбірливості було близьким до 6% у випадку з навушниками та до 10% для прослуховування через акустичні монітори. Для спотворення реверберацією відповідні значення стандартного відхилення були близькими до 8% та 12% відповідно.

Порівнюючи графіки (рис. 3.4, а і рис. 3.7, а), у випадку прослуховування сигналів через навушники ми бачимо, що значення розбірливості мови практично збігаються.

Зовсім інша ситуація спостерігається при порівнянні графіків (рис. 3.4, б і 3.8, а) для корпусних акустичних моніторів. Як видно, показані на рис. 3.8а значення розбірливості є значно нижчими, ніж на рис. 3.4б, і практично не відрізняються від результатів для випадку прослуховування сигналів через навушники (рис. 3.7, а).

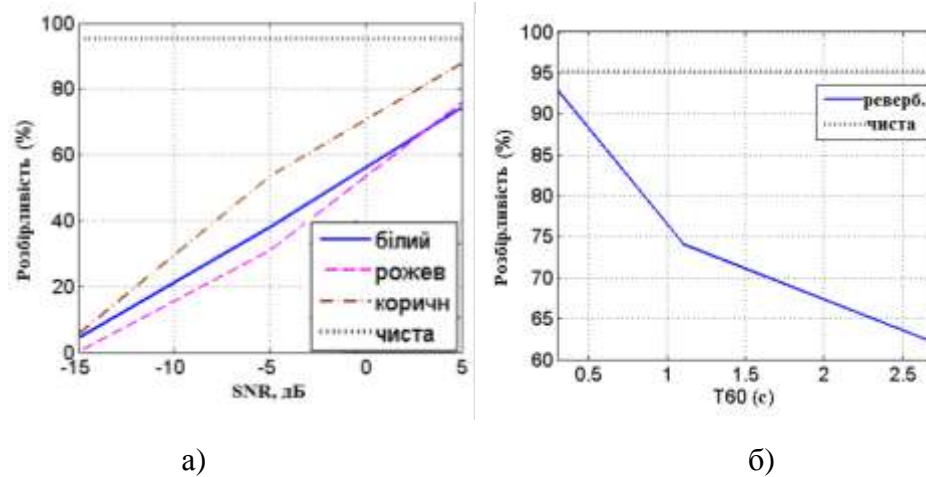


Рис. 3.7. Оцінки розбірливості при прослуховуванні через навушники для шуму (а) та реверберації (б)

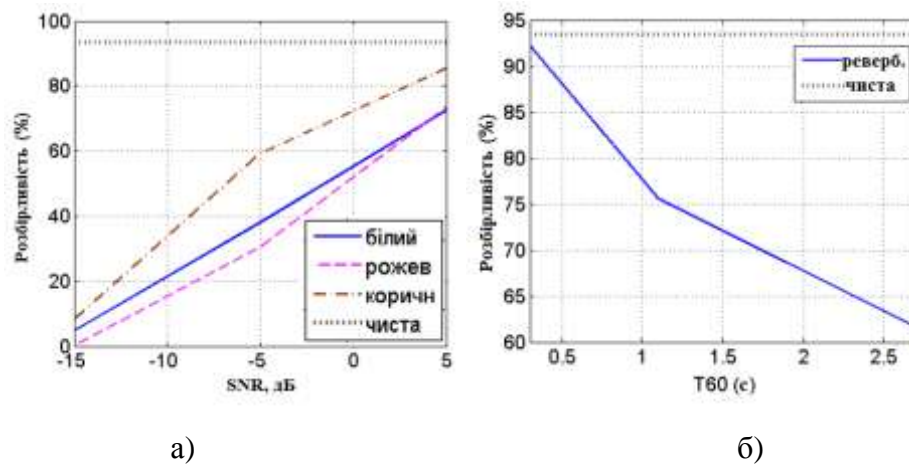


Рис. 3.8. Оцінки розбірливості при прослуховуванні через акустичні монітори для шуму (а) та реверберації (б)

Таким чином, можна зробити висновок, що проведена корекція програмного забезпечення та організації експерименту виявилася конструктивною. Отримані результати свідчать, по-перше, про те, що організація артикуляційних тестів та функцій програмного забезпечення може суттєво впливати на результати випробувань.

По-друге, дуже важливо, на наш погляд, зробити висновок про практичний збіг результатів прослуховування за допомогою акустичних моніторів та навушників. Слід зазначити, однак, що найімовірнішою причиною цього збігу було те, що слухачі зазвичай знаходяться на відстані 0,6-0,8 метрів від акустичних моніторів, де ефект реверберації незначний.

Загальним висновком може бути те, що прослуховування мовних сигналів, спотворених шумовими та ревербераційними взаємозв'язками, виконане із застосуванням запропонованої автоматизованої системи артикуляційних тестів, свідчить про ефективність та високу якість розробленої системи.

Перевірка явища аномально високої розбірливості мови під час прослуховування спотворених сигналів через акустичні системи показала, що найбільш імовірною причиною такої аномалії є особливості організації артикуляційних тестів та особливості програмного забезпечення.

3.2.3. Суб'єктивна оцінка впливу характеристик приміщення на розбірливість мови спотвореної шумом

Завданням експериментальних досліджень, описаних в даному підрозділі, було повторне дослідження проблеми комбінованого ефекту шуму, реверберації та розбірливості мови при бінауральному прослуховуванні на різних відстанях між диктором і слухачем.

Ще одна мета експерименту – спроба спростити організацію експериментальних досліджень. Дійсно, у [36] та [37] використовувалися 8-канальні та 29-канальні електроакустичні системи відповідно, а гучномовці було розміщено навколо слухача в заглушеній кімнаті. Крім того, розрахунок імпульсних відгуків приміщення був здійснений за допомогою ODEON та іншого дорогого програмного забезпечення. За допомогою методики, що була розроблена для даного експерименту, можуть бути проведені експериментальні

дослідження вдома, оскільки слухач може використовувати обладнання у вигляді комп'ютера та навушників, а також порівняно простої керуючої комп'ютерної програми [70].

Записи чистих сигналів робилися в заглушеному приміщенні кафедри акустики та акустoeлектроніки НТУУ «КПІ ім. І. Сікорського». Для записів використовувалися мікрофон Superlux ECM 999, зовнішня звукова карта PRESONUS AudioBox USB та програма Audacity. Запис виконувався із частотою дискретизації 44100 Гц та бітовою глибиною 16 біт. Формування тестових подразників здійснювалось у два етапи. Спочатку до чіткого мовного сигналу додався білий шум. Це підсумовування зважували, щоб забезпечити необхідне значення SNR в діапазон від мінус 15 дБ до плюс 5 дБ. Потім спотворену шумом мову відфільтровували двоканальним нерекурсивним фільтром за допомогою бінаурального імпульсного відгуку приміщення. Бінауральні імпульсні характеристики були виміряні на відстані 2,25-10,2 м від джерела звуку. Фільтрована мова прослуховувалася через навушники, ідентифікувалася та фіксувалася за допомогою клавіатура комп'ютера.

П'ятнадцять слухачів, що складаються з 11 чоловіків і 4 жінок з нормальним слухом, брали участь у тестуванні. Усі слухачі були у віці від 19 до 22 років (рідна мова – українська). Тестування проводилося вдома за допомогою комп'ютерів та спеціально розробленого програмного забезпечення. Сигнали були надані слухачам у випадковому порядку. Слухачі виконали набір тестів протягом 10-15 днів, щоб уникнути втоми.

Формування вихідного сигналу виконано за допомогою згортки імпульсної характеристики приміщення $h(t)$ та спотвореного шумом мовного сигналу:

$$x(t) = [s(t) + n(t)] \otimes h(t)$$

Для досліджень використовували базу даних [19,20] шести бінауральних імпульсних відгуків. Імпульсні відгуки належать до лекційної аудиторії 10,8 x

10,9 x 3,15 м, яка має 3 скляних вікна, бетонні стіни, паркет, дерев'яні столи та стільці. Гучномовець розміщувався на лекційній трибуні, а мікрофон розміщувався за різними столами зі збільшенням відстані d до динаміка. Часи реверберації RT_{60} для вимірюваних позицій показані в таблиці 3.1.

Таблиця 3.1 Час реверберації

d (м)	2.25	4.00	5.56	7.10	8.68	10.2
RT_{60} (с)	0.70	0.72	0.79	0.80	0.81	0.83

Форми пари бінауральних імпульсних характеристик на відстані 2,25 м і 10,2 м наведено на рис. 3.9а та 3.9б.

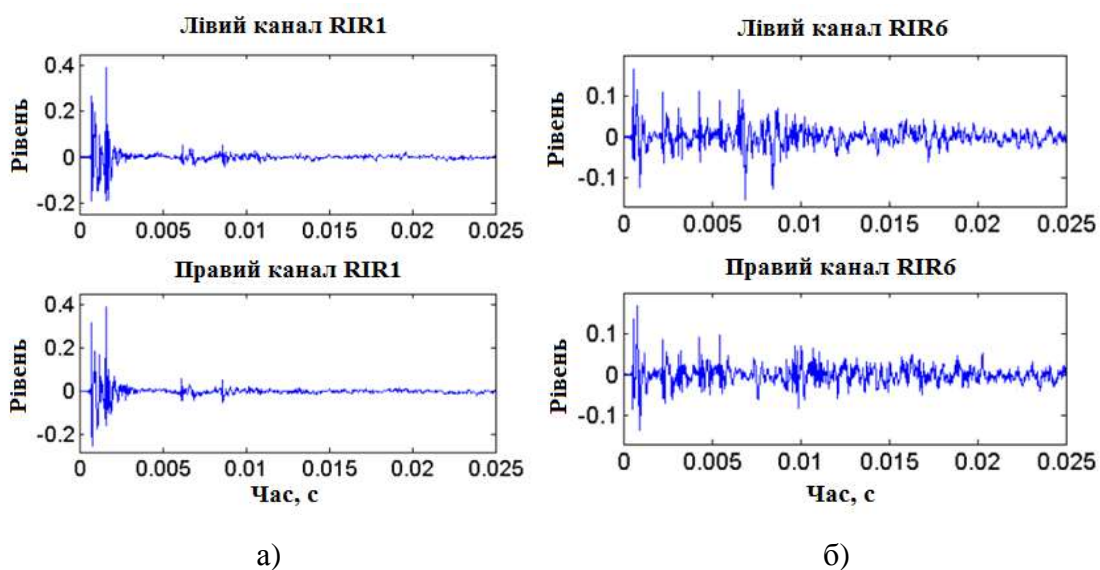


Рис. 3.9 Форми пари бінауральних імпульсних відгуків на відстані 2,25 м (а) і 10,2 м (б)

Значення співвідношення енергії раннього та пізнього часу (C_{50}) для всіх шести імпульсних відгуків 2 показано на рис. 3.10. З рис. 3.10 випливає, що всі шість використаних імпульсних відгуків можуть забезпечувати хорошу розбірливість мови. Дійсно, відомо, що розбірливість складу не нижче 80% (що відповідає розбірливості фраз не нижче 95%) досягається за умови $C_{50} > -2$ дБ [41]. Крім того, з рис. 2 випливає, що значення індексу C_{50} змінюються дуже

мало на відстані 5-10 м, тобто вплив ранніх перевідбиттів на розбірливість мови для них майже однаковий.

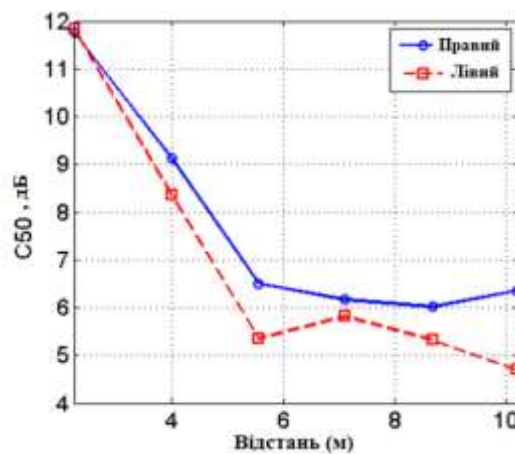


Рис. 3.10. Значення C_{50} для імпульсних відгуків кімнат

Усереднені результати артикуляційних тестів для 15 слухачів показані на рис. 3.11а, а оцінки стандартних відхилень цих результатів наведені на рис. 3.11б. Як видно, для малих значень SNR ($SNR = -10 \dots -5$ дБ) розбірливість мови на відстані 7-10 м була приблизно на 10% вище, ніж на відстані 2,25 м. Для середніх та високих значень SNR ($SNR > 3-5$ дБ) розбірливість мови на відстані 4-10 м виявилася меншою за одиницю на відстані 2,25 м. Таким чином, ефект приміщення як достатньо ефективного фільтра для подавлення шуму був найбільш вираженим, коли поєднувалися два фактори: невелике значення SNR та складна структура ранніх відбиттів. При середніх і великих значеннях SNR шкідлива дія пізніх відбиттів переважає над корисним ефектом ранніх відбиттів. До речі, слід бути обережним, говорячи про корисність ранніх відбиттів, враховуючи їх руйнівний вплив на мовленнєвий спектр, хоча цей ефект частково компенсується бінауральним прослуховуванням [42].

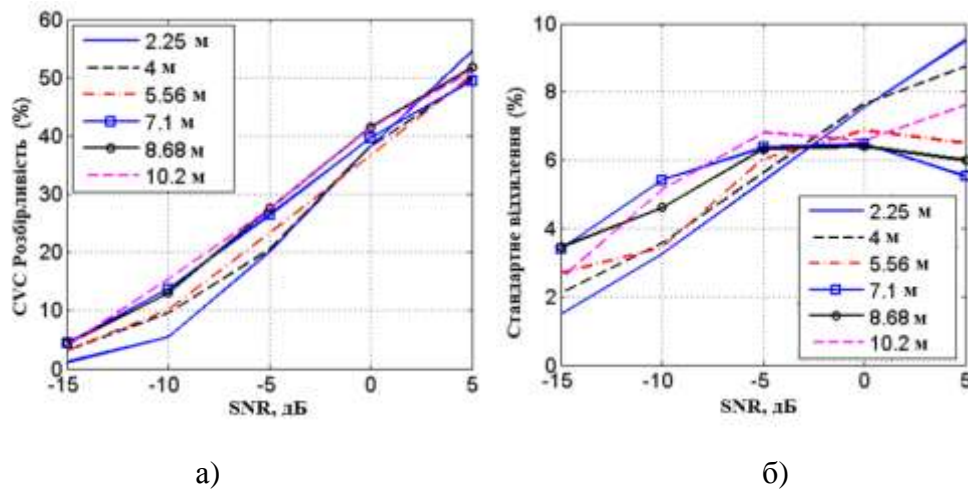


Рис. 3.11. Усереднені оцінки розбірливості мови (а) та стандартні відхилення (б), на різні відстані d

3.3. Оцінювання розбірливості мови через оцінювання якості мови.

Зазначимо, що при оцінці розбірливості та якості мовлення застосовуються різні кількісні міри, оскільки якість мовлення та його розбірливість не є тотожними поняттями, хоча ці поняття тісно пов'язані. У справедливості цього твердження легко впевнитися, порівнявши шкали суб'єктивного оцінювання якості та розбірливості мови. Дійсно, при оцінюванні якості мови фактично оцінюється рівень негативних емоцій слухача. Проте, коли оцінюється розбірливість мови, оцінюється кількість отриманої інформації шляхом підрахування відносної кількості правильно почутих елементів мови (склади, слова, фрази).

Даний підрозділ присвячено встановленню принципової можливості оцінювати розбірливість мови через оцінювання якості мови. Така можливість встановлюється шляхом порівняння оцінок якості мовних сигналів з оцінками розбірливості мови, спотвореної білим, рожевим і коричневим шумом різної інтенсивності. Важливо зазначити, що постановка даної задачі спричинена значною трудомісткістю оцінювання розбірливості не тільки суб'єктивними, але й навіть об'єктивними (інструментальними) методами.

У даній дисертаційній роботі оцінювання розбірливості мови реалізовано суб'єктивним методом, шляхом оцінювання розбірливості складів. Оцінювання якості мови реалізовано як суб'єктивним методом, за шкалою DMOS, так і об'єктивним методом, із використанням мір SSNR, LSD, fwSNR, BSD та WB-PESQ.

Зауважимо, що у деяких роботах представлені результати досліджень, спрямованих на оцінку розбірливості мови із використанням мір якості мови [31], [32], [33]. Взагалі, спроби використати «чужі» показники, як правило, пояснюються низькою ефективністю або відсутністю обґрунтованості "рідних" показників. Прикладом такої спроби є припущення про використання вимірювання якості PESQ для оцінки розбірливості мови в каналах зв'язку з низькошвидкісними кодеками [32].

Іншим прикладом є оцінювання розбірливості мови на виході слухового апарату при наявності фонових завад у формі розмови між двома людьми [31]. У якості тестових сигналів було використано записи ста речень, при цьому оцінювалася розбірливість як фраз, так і складів. Об'єктивне оцінювання якості мови проводилося за допомогою широкосмугової версії PESQ (WB-PESQ). Значення коефіцієнта кореляції r між оцінками PESQ та оцінками суб'єктивної розбірливості виявилися дуже високими: $r = 0,91, \dots, 0,99$ для фраз, та $r = 0,86, \dots, 0,95$ для складів.

Очевидно, вивчення інших, більш простих для розрахунків мір, ніж PESQ, також представляє значний інтерес. На жаль, такі альтернативні міри не було розглянуто в [31] та [32]. Цей недолік в якійсь мірі було усунено в роботі [33], де були представлені результати досліджень досить великого переліку об'єктивних мір [33] щодо їх придатності у ролі мір достовірності розпізнавання мовлення. При цьому, крім вузькосмугового вимірювання PESQ (NB-PESQ), було розглянуто такі міри як SSNR, fwSNR, LLR, IS, CEP, WSS. При оцінці розбірливості приголосних звуків і розбірливості речень, тільки міри PESQ ($r = 0.77 \dots 0.79$) та fwSNR ($r = 0.68 \dots 0.81$) виявилися помірно ефективними.

Цікаво, що міри CEP, LLR та SSNR виявилися не дуже ефективними, а показники IS та WSS виявилися найменш ефективними. На жаль, міри LSD та BSD, які легко зрозуміти та обчислити, не було розглянуто в [33]. Недоліком є також відсутність результатів досліджень дії "стандартного" білого, рожевого та коричневого шуму на розбірливість мови. Нарешті, спірним є використання коефіцієнта кореляції як показника об'єктивної надійності мір якості.

У даній дисертаційній роботі зазначені недоліки усунено шляхом проведення низки відповідних експериментальних досліджень.

3.3.1. Організація експериментальних досліджень

Експериментальні дослідження склалися із 3-х частин.

У першій частині досліджень суб'єктивним (артикуляційним) методом оцінювалася залежність розбірливості мови від відношення сигнал-шум. З цією метою 20 молодих людей без вад слуху (середній вік – 22 роки) виконували оцінку розбірливості мови односкладових звукосполучень типу «приголосний-голосний-приголосний» (consonant–vowel–consonant – CVC), спотвореного білим, рожевим і коричневим шумом. Оцінка проводилася за допомогою спеціально розробленої комп'ютерної програми [70]. Слухачі прослуховували через навушники 3 таблиці, кожна з яких містила 50 складів. Односкладові слова були відтворені у випадковому порядку.

У другій частині експерименту суб'єктивно оцінювалася залежність якості мовних сигналів за шкалою DMOS [30] від відношення сигнал-шум. Оцінювання виконувалося іншими 10 слухачами середнього віку 22 років, без вад слуху, при цьому слухачі прослуховували записи спотворених сигналів у випадковому порядку.

У третій частині експериментальних досліджень обчислювалися значення залежностей об'єктивних мір якості мовних сигналів, таких як SSNR, LSD, BSD, fwSNR та PESQ, від відношення сигнал-шум.

Порівняння отриманих залежностей дозволяє зробити певні висновки щодо принципової можливості оцінювання розбірливості мови через оцінювання її якості.

3.3.2. Результати суб'єктивного оцінювання розбірливості та якості мовних сигналів

Отримані артикуляційним методом (20 слухачів, кожен з яких прослуховував 3 таблиці з 50 складів типу «приголосний-голосний-приголосний») усереднені суб'єктивні оцінки розбірливості мови як функції відношення сигнал-шум показані на рис. 3.12а.

Залежності суб'єктивних оцінок якості мовлення від відношення сигнал-шум для безперервної мови (8 дикторів – 4 чоловіки та 4 жінки, 10 слухачів), спотвореної білим, рожевим і коричневим шумом, наведено на рис. 3.12б.

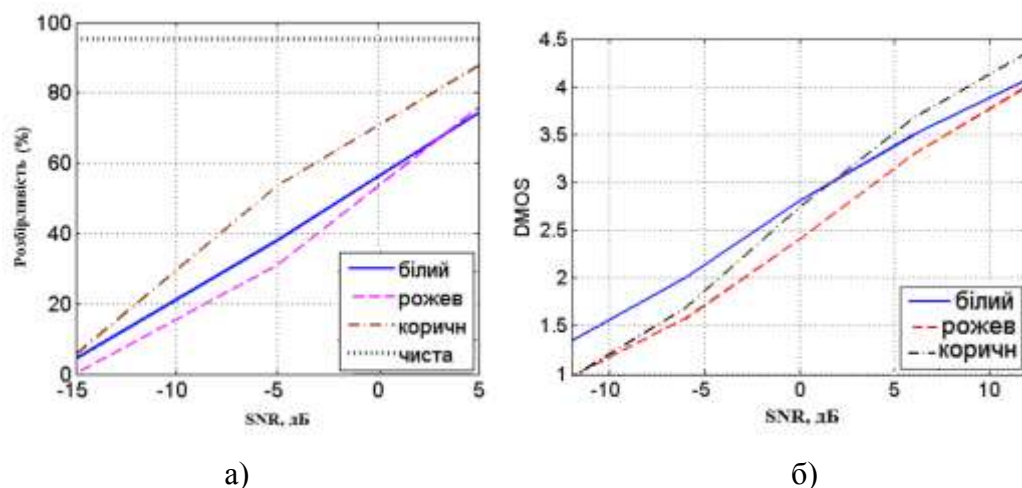


Рис. 3.12. Усереднені оцінки розбірливості мови (а) і якості мови (б).

Як видно з рис. 3.12а, розбірливість мови є найгіршою при дії рожевого шуму, хоча існує тенденція до наближення результатів для білого та рожевого шумів при $SNR > 5$ дБ. У той же час, розбірливість мови виявилася найвищою при дії коричневого шуму.

Порівнюючи графіки на рис. 3.12а та 3.12б, можна зробити висновок, що поведінка графіків є подібною для середніх і великих значень SNR ($SNR > 0$ дБ). Проте ця подібність втрачається у випадку малих значень SNR ($SNR < 0$ дБ).

3.3.3. Порівняння результатів об'єктивного оцінювання якості мовних сигналів із результатами суб'єктивного оцінювання їх якості та розбірливості

Розглянемо тепер результати оцінювання якості мови із застосуванням об'єктивних мір якості. Оцінки міри SSNR показані на рис. 3.13.

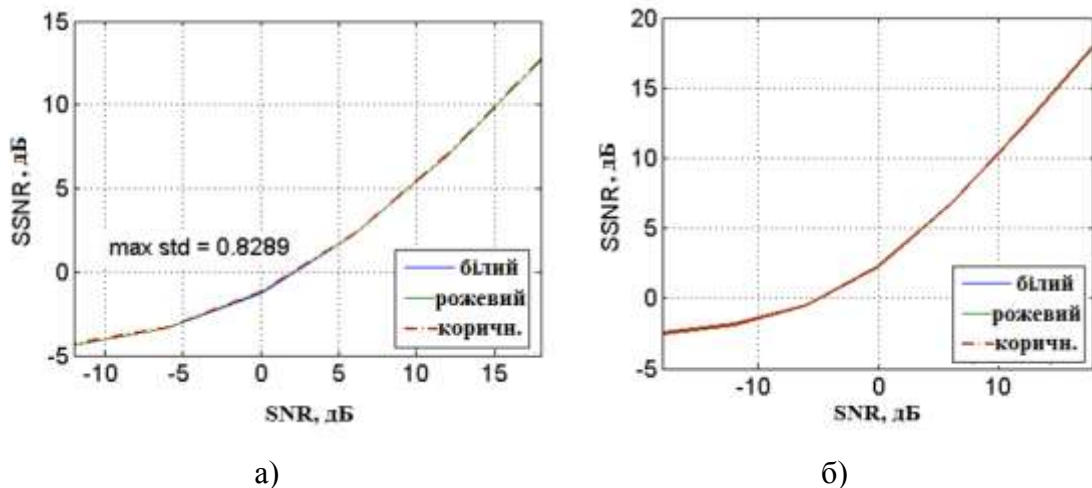


Рис. 3.13. Оцінки міри SSNR для мови (а) і коричневого шуму (б)

Реакція значення SSNR на SNR була очікуваною, оскільки якість сигналу, спотвореного шумом, повинна монотонно збільшуватися при збільшенні SNR. Проте міра SSNR є єдиною мірою серед розглянутих, для яких графіки не залежать від спектральних властивостей шуму. Ця поведінка пояснюється змістом міри (1).

Дійсно, ми б мали $SSNR = SNR$ якби суміш $y(n) = x(n) + \xi(n)$ було використано в знаменнику (1.1), оскільки оцінки коефіцієнтів дисперсії $\tilde{D}_x / \tilde{D}_\xi$ сигналу $x(n)$ та шуму $\xi(n)$ розраховуються для кожного з m -х сегментів.

У цьому випадку $SSNR$ не повинно залежати від спектрів процесів $x(n)$ та $y(n)$. Однак, оскільки нас цікавить різниця у формі сигналів $x(n)$ та $y(n)$, вони повинні бути нормалізовані за своїми стандартними відхиленнями. Отже, рівняння $SSNR \approx SNR$ виконується тільки для $SNR \geq 10$ дБ і тільки коли сигнал $x(n)$ є стаціонарним. Це чітко підтверджується графіком рис. 3.13б, для сигналу у вигляді стаціонарного коричневого шуму.

Порівнюючи рис. 3.13а та 3.13б ми бачимо, що рівняння $SSNR \approx SNR$ не виконується у випадку нестаціонарного мовного сигналу навіть для $SNR \geq 10$ дБ. Це вказує на те, що нестаціонарний характер мовного сигналу призводить до нерівності $SSNR < SNR$ для $SNR \geq 0$ дБ. Для значення $SNR < -2$ дБ, ситуація на рис. 3.13а змінюється і нерівність $SSNR > SNR$ виконується. Це пояснюється тим, що головну роль відіграє невідповідність чистої та зашумленої форми сигналів для малих значень SNR . Дійсно, у випадку стаціонарного випадкового сигналу (рис. 3.13, б) ми бачимо, що нерівність $SSNR > SNR$ стає помітно вираженою при $SNR \leq 5$ дБ. Таким чином, інтервал $-2 \leq SNR \leq 5$ дБ обмежує зони впливу на поведінку залежності $SSNR(SNR)$ для таких "конкуруючих" факторів як нестаціонарний характер мовного сигналу та відносний рівень шуму.

Графіки для мір LSD та BSD (рис. 3.14, а та 3.14, б) дуже схожі.

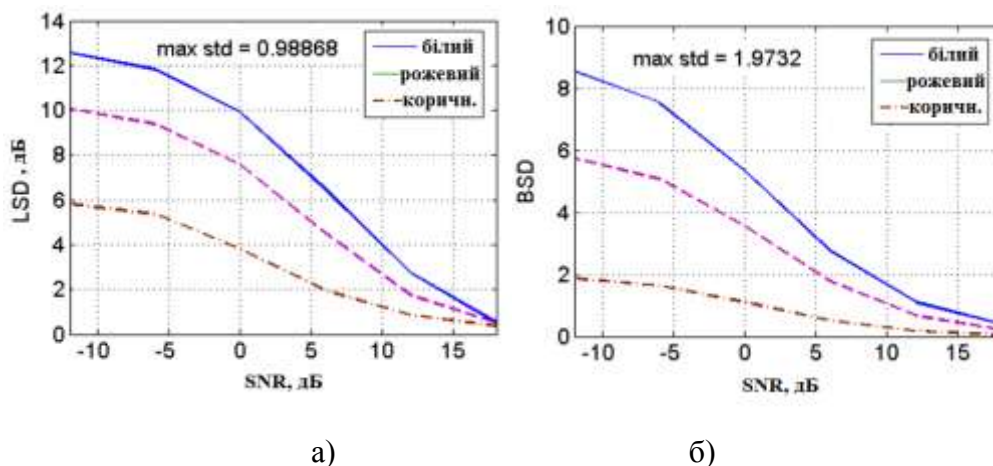


Рис. 3.14. Оцінювання мір LSD (а) та BSD (б)

Виходячи з цих графіків, можна зробити висновок, що коричневий шум найменше погіршує якість сигналу, тоді як білий шум є найбільш руйнівним у всьому діапазоні визначених значень SNR. Порівняння цих графіків з результатами суб'єктивної оцінки якості мовлення та розпізнавання мови (рис. 3.12) показує, що узгодження не є повним.

Поведінка залежності $fwSNR(SNR)$ (рис. 3.15, а) є досить суперечливою для різних інтервалів параметра SNR. Однак існує узгодження із суб'єктивними оцінками якості мови та розбірливості для $SNR > 10$ дБ, а саме: коричневий шум оцінюється як найменш руйнівний, а рожевий як найбільш руйнівний.

Поведінка графіків PESQ (рис. 3.15, б) добре співпадає з суб'єктивними оцінками якості мовлення та розбірливості при $SNR > 0$ дБ (рис. 3.12). У той же час міра PESQ не реагує адекватно на ситуацію сильного спотворення сигналу адитивним шумом, а саме: при $SNR < -5$ дБ значення PESQ зменшуються зі збільшенням SNR для коричневого шуму.

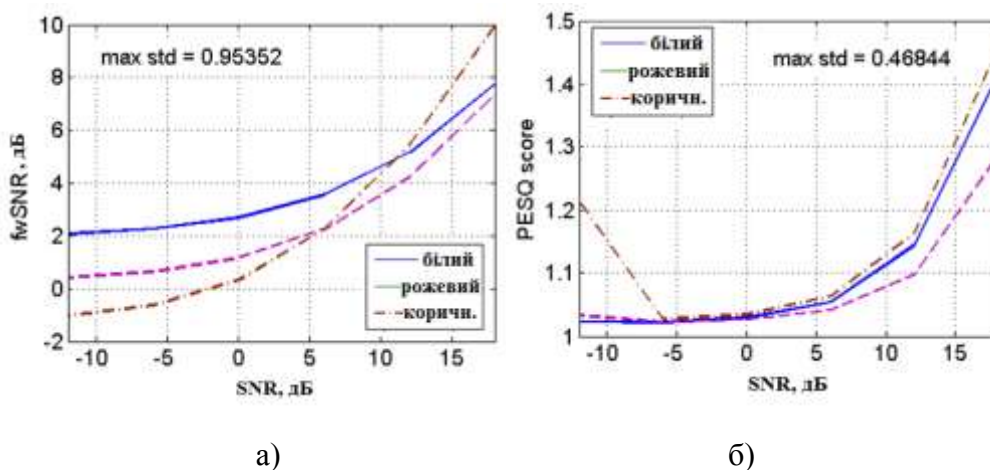


Рис. 3.15 Оцінювання мір: $fwSNR$ (а), PESQ(б)

3.3.4. Висновки за отриманими результатами

Таким чином, оцінювання п'яти об'єктивних показників якості мови, таких як SSNR, LSD, $fwSNR$, BSD та PESQ для мовних сигналів, спотворених шумом, дозволило виявити деякі особливості, властиві цим показникам якості.

Виявилося, що показник SSNR, на відміну від інших мір, є нечутливим до форми спектральної щільності шумової потужності, що погано узгоджується з поведінкою суб'єктивних мір розбірливості мови. Поведінка мір LSD та BSD лише частково узгоджується з результатами суб'єктивної оцінки якості та розбірливості зашумленої мови, а саме: коричневий шум правильно оцінений, як такий, що призводить до найменшого погіршення розбірливості мови.

Поведінка міри PESQ добре узгоджується з суб'єктивними оцінками при $SNR > 0$ дБ, тоді як для міри fwSNR це узгодження спостерігається лише при $SNR > 10$ дБ.

Порівняння оцінок суб'єктивної якості та розбірливості мови, спотвореної білим, рожевим і коричневим шумом, вказує на те, що результати є узгодженими для $SNR > 0$ дБ.

Отримані результати вказують, з одного боку, на принципову можливість заміни суб'єктивного оцінювання розбірливості зашумленої мови об'єктивною оцінкою якості мови. У той же час слід враховувати наявність досить серйозних обмежень щодо впровадження такого підходу.

Висновки до розділу 3

1. Вперше розроблено автоматизовану систему артикуляційних випробувань каналів передачі української мови. За результатами випробувань розроблену автоматизовану систему артикуляційних випробувань слід визнати дієвою. Запропоновано рекомендації із подальшого покращення розробленої системи, яку можна використовувати в якості прототипу при створенні національних стандартів оцінювання розбірливості мови в каналах зв'язку та в приміщеннях.
2. Експериментально досліджено дієвість розробленої автоматизованої системи артикуляційних випробувань та достовірність результатів.

Одержані результати артикуляційних випробувань, отримані з застосуванням запропонованих артикуляційних таблиць та програмного комплексу, задовільно узгоджуються із результатами, наведеними в літературних джерелах, що свідчить про дієвість та досить високу ефективність таблиць та програмного забезпечення.

3. Результати експериментальних досліджень засвідчили можливість суттєвого підвищення розбірливості мови при прослуховуванні спотворених записів через акустичні монітори, а не через навушники, що можна пояснити дією таких чинників як наявність потужних ранніх відбиттів в невеликих приміщеннях, використання двох джерел прямого звуку замість одного, а також бінауральне сприйняття звуку. Крім того, встановлено, що суттєвим чинником може бути недовершеність організації експериментальних досліджень, наприклад, надмірний об'єм інформації, що надається слухачеві на моніторі комп'ютера, або нераціональний порядок прослуховування тестових сигналів.
4. Уточнено межі застосування об'єктивних оцінок якості мови замість суб'єктивних оцінок розбірливості зашумленої мови, що дозволяє, за певних значень відношення сигнал-шум ($SNR > 0$ дБ), значно спростити процедуру об'єктивного оцінювання розбірливості мови.

РОЗДІЛ 4. ЗІСТАВЛЕННЯ АЛГОРИТМІВ ОБРОБКИ МУЗИЧНИХ СИГНАЛІВ У БАГАТОМІКРОФОННИХ МАСИВАХ

Наведені в роботі [57] результати об'єктивного оцінювання якості мовних сигналів свідчать, що класичний метод формування характеристики спрямованості DS дав змогу отримати більш якісні результати часово-просторової обробки мовних сигналів, у порівнянні з іншими методами (алгоритми SDA, MVDR, LCMV). Однак, на жаль, в [57] відсутня спроба пояснити отримані результати об'єктивного оцінювання. Крім того, відсутні результати суб'єктивного оцінювання якості музичних сигналів на виході алгоритмів просторово-часової обробки, що є суттєвим недоліком з огляду на еталонний характер результатів суб'єктивного оцінювання якості акустичних сигналів.

У даному розділі вказані недоліки усунено шляхом проведення суб'єктивного та об'єктивного оцінювання якості музичного сигналу, спотвореного шумом. При цьому виконано порівняння якості сигналів, отриманих двома методами просторово-часової фільтрації, а саме – класичного методу DS та методу LCMV, який забезпечує незмінну у частотному діапазоні характеристику спрямованості. Алгоритми SDA та MVDR не було залучено до досліджень, оскільки, по-перше, вони належать до того ж класу алгоритмів, що й алгоритм LCMV, а по-друге, це призвело б до неприйнятного збільшення часу суб'єктивного оцінювання.

4.1. Комп'ютерна модель багатомікрофонного масиву

При експериментальних дослідженнях в даній роботі використано комп'ютерні моделі просторово-часової обробки за алгоритмами DS і LCMV.

Схема просторово-часової обробки музичного сигналу дискретною лінійною антеною з N приймачів зображена на рис. 4.1.

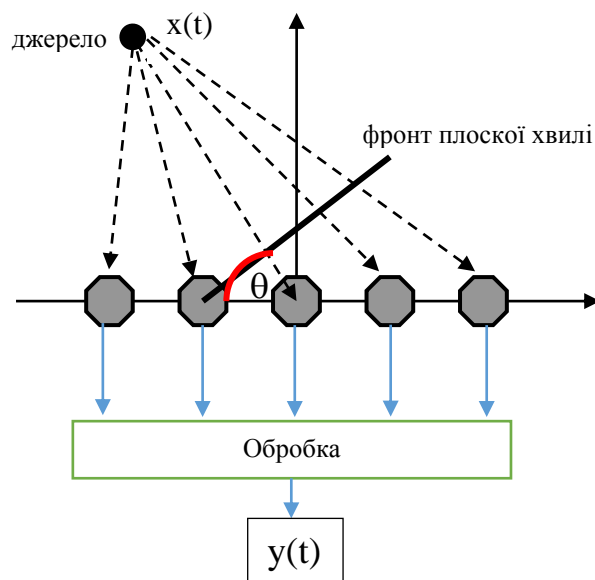


Рис. 4.1. Узагальнена блок-схема формування характеристики спрямованості

У блоці обробки виконується додавання сигналів з виходів окремих мікрофонів з певними затримками (у випадку використання методу затримок та сум) та спеціальним чином обчисленими ваговими коефіцієнтами (у випадку з методом LCMV). Обробка сигналу проводиться в частотній області. На виході алгоритму обробки, після виконання оберненого перетворення Фур'є, отримаємо сигнал $y(t)$ (рис. 4.2), якість якого потім піддається оцінюванню.

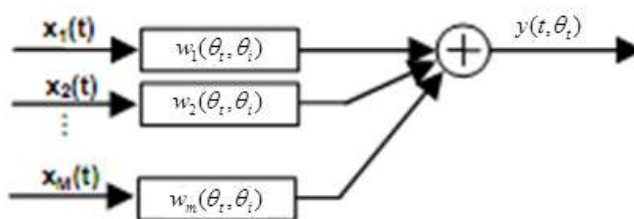


Рис. 4.2. Блок-схема обробки сигналу [44]

Рівняння для функції вагових коефіцієнтів методу LCMV має вигляд [57]:

$$w_m(\theta_t, \theta_i) = (R_{xx} + \varepsilon I)^{-1} C \left[C^H (R_{xx} + \varepsilon I)^{-1} C \right]^{-1} C^H w \quad (4.1)$$

де $R_{xx} = E[xx^H]$ – визначає матрицю просторової кореляції, а $E[\cdot]$ – оператор математичного очікування, I – одинична матриця розміром $M \times M$, ε – параметр регуляризації, що залежить від частоти, C – матриця обмежень.

При комп'ютерному моделюванні в якості прототипу було обрано комп'ютерну програму `MicrophoneBeamformerExample.m` зі стандартної бібліотеки Matlab.

На рис. 4.3 наведено характеристики спрямованості в залежності від частоти, в [42] їх називають картами спрямованості.

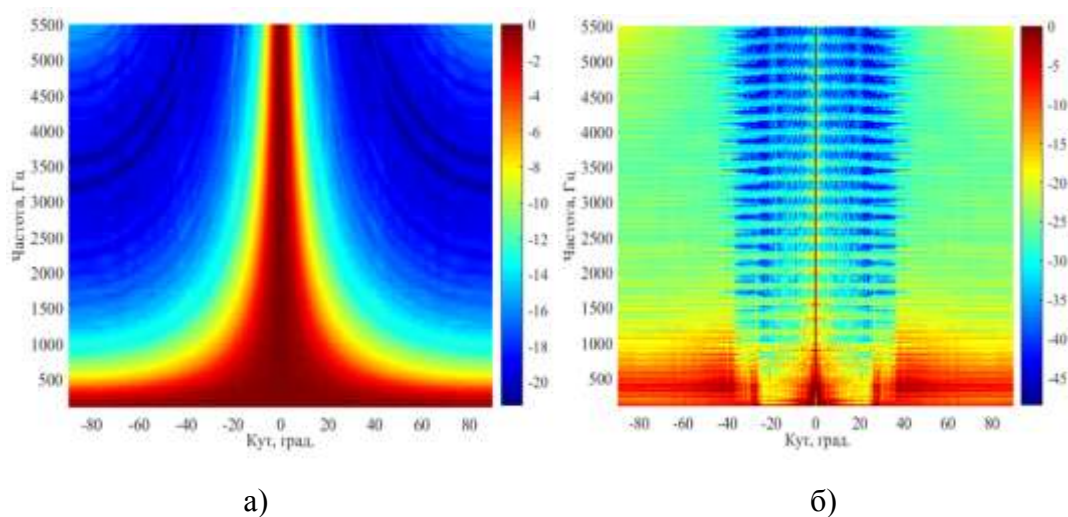


Рис. 4.3. Карти спрямованості для алгоритмів DS (а) та LCMV (б)

Згідно рис. 4.3а, в антени DS в смузі частот від 100 Гц приблизно до 300 Гц розкрив головного максимуму є близьким до 15 градусів. На рис. 4.3б показано карту спрямованості LCMV антени, яка має бічні максимуми в смузі частот 100-500 Гц, та незалежний від частоти головний максимум, з розкритом у приблизно 1 градус.

4.2. Суб'єктивне оцінювання якості музики із застосуванням мікрофонних масивів

Для експериментальних досліджень було використано наступні параметри мікрофонної решітки:

- антена – лінійна, дискретна;
- кількість мікрофонів $N=10$, мікрофони розташовані з кроком 0.05 м;
- кожен мікрофон має АЧХ прямокутної форми з граничними частотами 20 Гц та 20 кГц;

При цьому розглядаються чотири ситуації розташування джерела шуму:

- перша, друга і третя ситуації – одиночне джерело шуму у вигляді стола зі співрозмовниками, розташованого під кутами до джерела музичного сигналу 15, 45 та 90 градусів, відповідно;
- четверта ситуація – кілька джерел шуму у вигляді двох столів зі співрозмовниками, розташованих під кутами до джерела музичного сигналу 45 та 120 градусів.

Розглянуті ситуації обрано як такі, що є типовими для розважальних закладів.

У якості сигналу використано соло-партію саксофону, яку записано в студійних умовах, без впливу реверберації та шуму, а в якості шуму використано запис розмови та сміху відвідувачів ресторану, зроблений в реальних умовах. При суб'єктивному оцінюванні слухачі у випадковому порядку виконували прослуховування для трьох співвідношень сигнал/шум (-10 дБ, 0 дБ та 10 дБ), щоб оцінити ефективність фільтрації в різних завадових умовах. При цьому прослуховувалися 4 види сигналів:

- 1) еталонний сигнал (запис інструменту, який необхідно виділити з суміші);
- 2) суміш сигналу та шуму на виході одного з мікрофонів масиву (тип обробки 2);
- 3) суміш сигналу та шуму на виході DS-антени (тип обробки 3);
- 4) суміш сигналу та шуму на виході LCMV-антени (при 4-х різних значеннях параметру пост-фільтрації: порядок нерекурсивного фільтра 5, 10, 20 та 40) (типи обробки 4-7).

Якість музичного сигналу, підданого просторовій обробці в багатомікрофонних системах, оцінювалася за допомогою кількох шкал. За 5-бальною системою за шкалою DMOS оцінювалась якість музичного сигналу, що давало змогу оцінити ступінь спотворення музичного сигналу на виході багатомікрофонних систем. Крім того, за 5-бальними шкалами також оцінювалися рівень завади (оцінка ступеню придушення шуму по відношенню до музичного сигналу), та загальне враження слухача від почутого (інтегральний показник, який включає в себе перші два показники).

4.2.1. Пілотні експериментальні дослідження з оцінювання якості музичних сигналів у багатомікрофонних системах

Усереднені результати експериментів наведено на рисунках 4.4-4.7. Кількість слухачів становила 21 людину, 20 з яких – віком від 20 до 27 років та один 70 років. Усі слухачі – зі здоровим слухом.

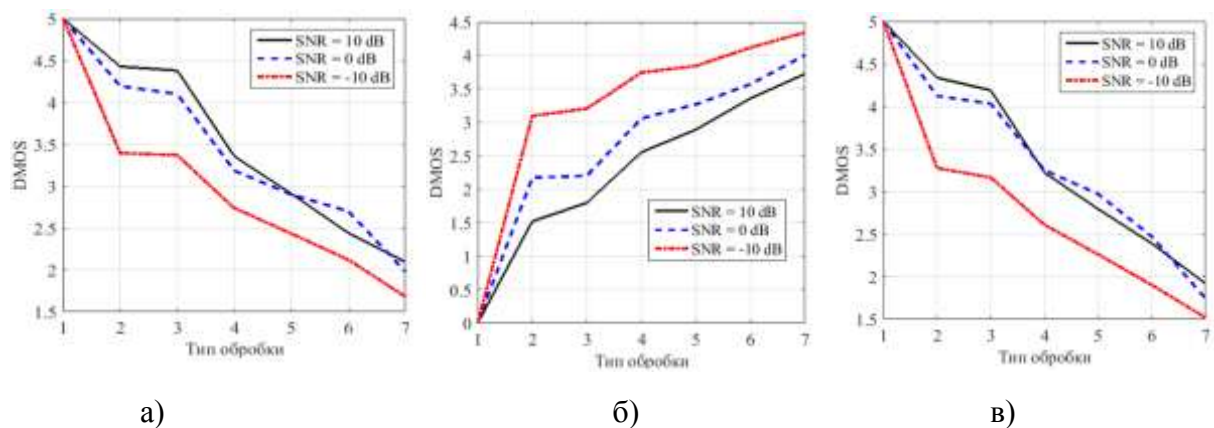


Рис. 4.4. Результати суб'єктивного тесту для ситуації 1: оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

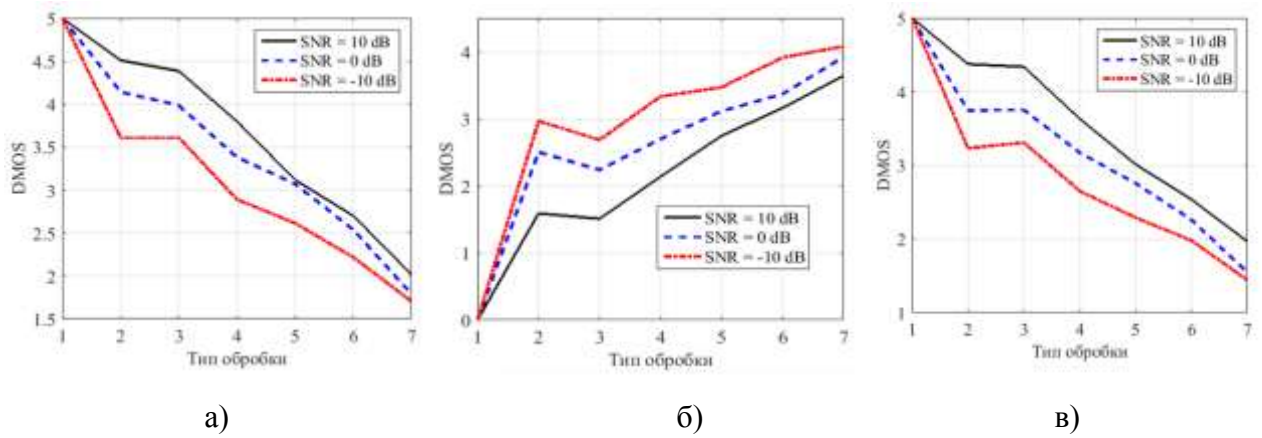


Рис. 4.5. Результати суб'єктивного тесту для ситуації 2: оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

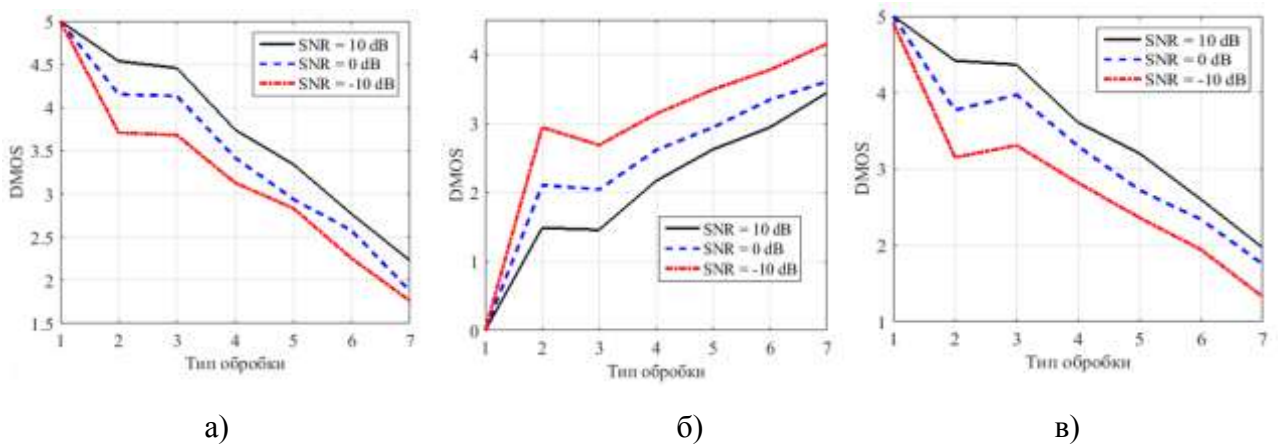


Рис. 4.6. Результати суб'єктивного тесту для ситуації 3: оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

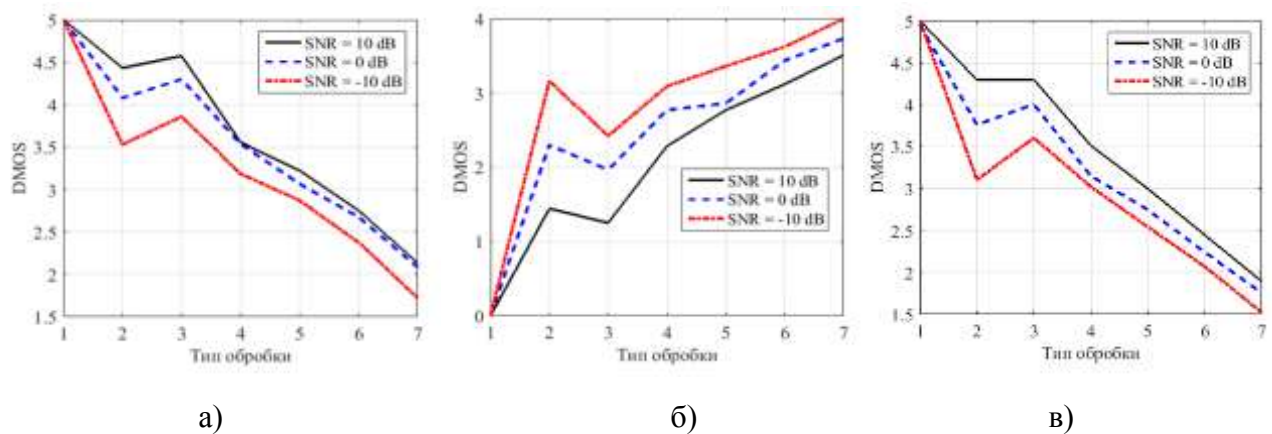


Рис. 4.7. Результати суб'єктивного тесту для ситуації 4: оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

У результаті суб'єктивних досліджень якості музичного сигналу виявилось, що найвища якість, та, відповідно, найнижчий рівень завад можна отримати, застосовуючи алгоритм DS. При цьому викликає подив підозріло значне зниження якості сигналу при застосуванні алгоритму LCMV – якість сигналу в цьому випадку є навіть нижчою, ніж при прослуховуванні сигналу з одиночного мікрофона.

Об'єктивне оцінювання здійснено в середовищі Matlab із використанням чотирьох мір якості, таких як SSNR, LSD, BSD та PEAQ. Результати оцінювання об'єктивних показників якості наведено на рис. 4.8-4.11.

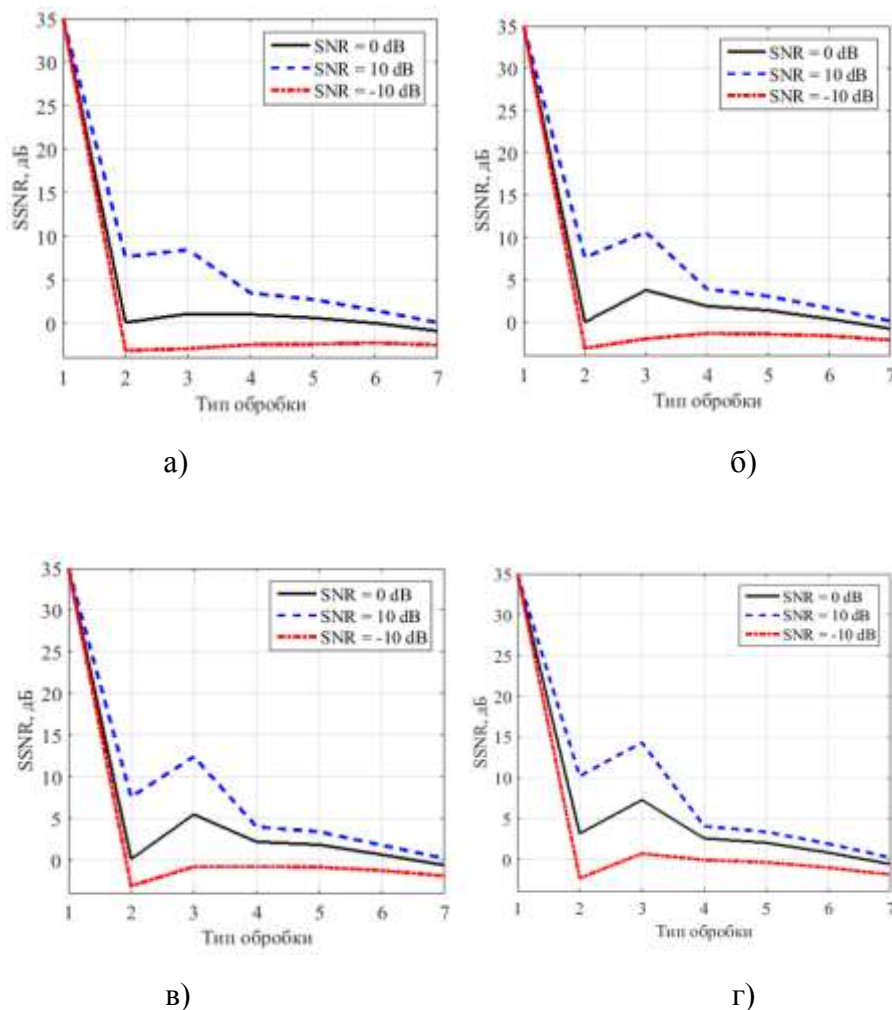


Рис. 4.8. Міра якості SSNR: ситуація 1(а), ситуація 2(б), ситуація 3(в), ситуація 4(г).

Різниця показника SSNR (рис. 4.8) для типів обробки DS та LCMV становить від 8 до 15 дБ, причому для алгоритму LCMV максимальне значення SSNR відповідає ситуації 4, тобто постфільтрації з мінімальним порядком фільтрів, що дорівнює 5 – (рис. 4.8, г).

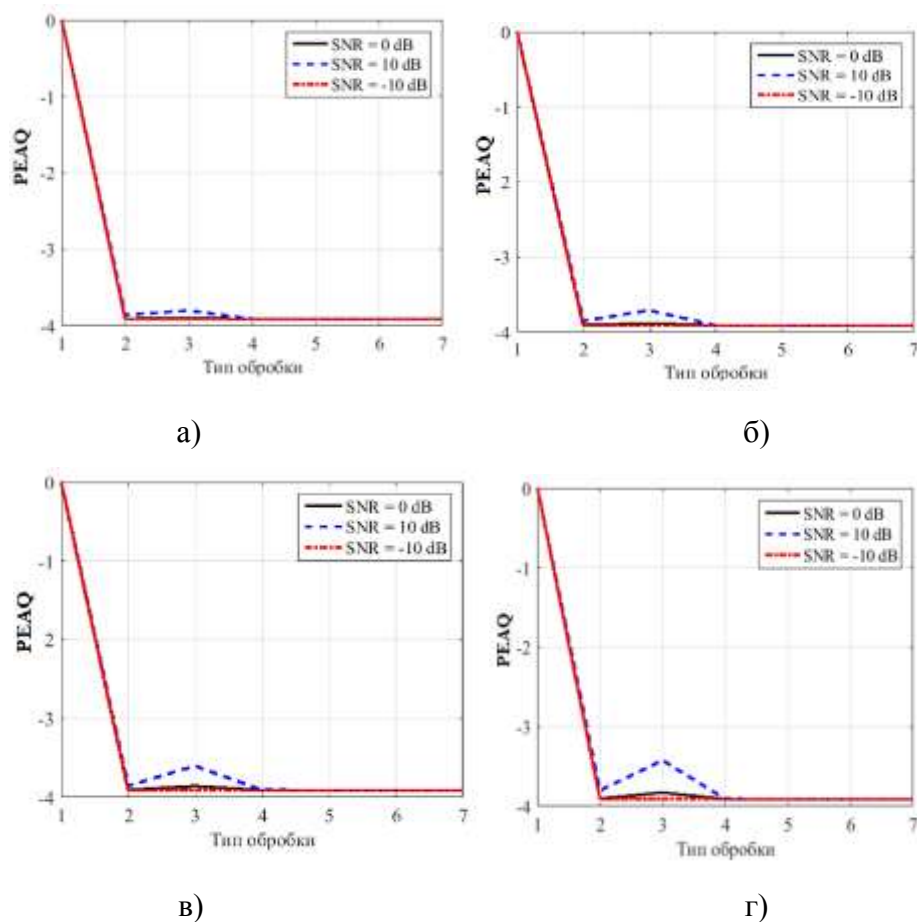
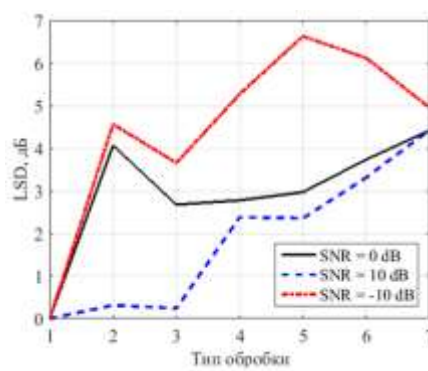
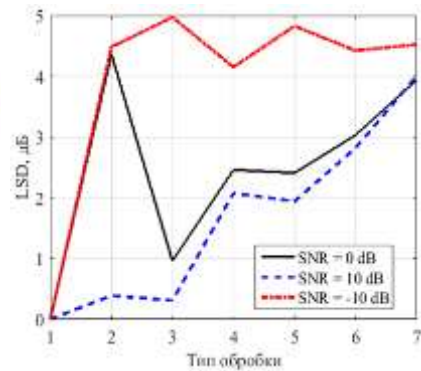


Рис. 4.9. Міра якості PEAQ: ситуація 1(а), ситуація 2(б), ситуація 3(в), ситуація 4(г).

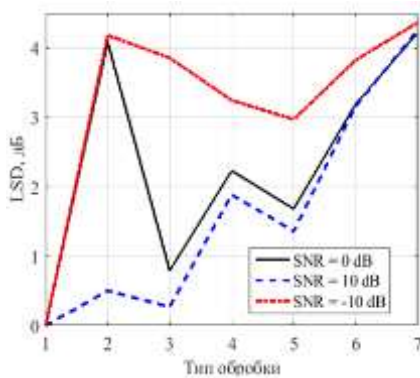
Значення показника якості звукового сигналу PEAQ загалом відповідають результатам суб'єктивних досліджень. Слід зазначити, що при розробці міри PEAQ малося на увазі її застосування до музичних стерео сигналів із дуже малим ступенем спотворення.



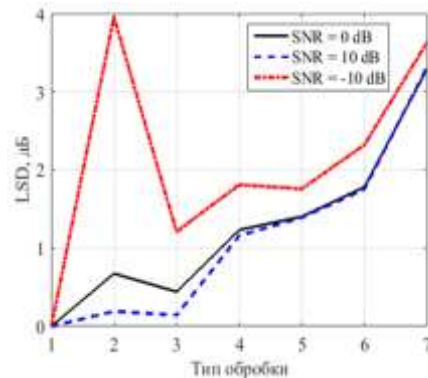
а)



б)



в)



г)

Рис. 4.10. Міра якості LSD: ситуація 1(а), ситуація 2(б), ситуація 3(в), ситуація 4(г).

Поведінка показників LSD (рис 4.10) та BSD (рис 4.11) є аналогічною до поведінки показників SSNR та PEAQ.

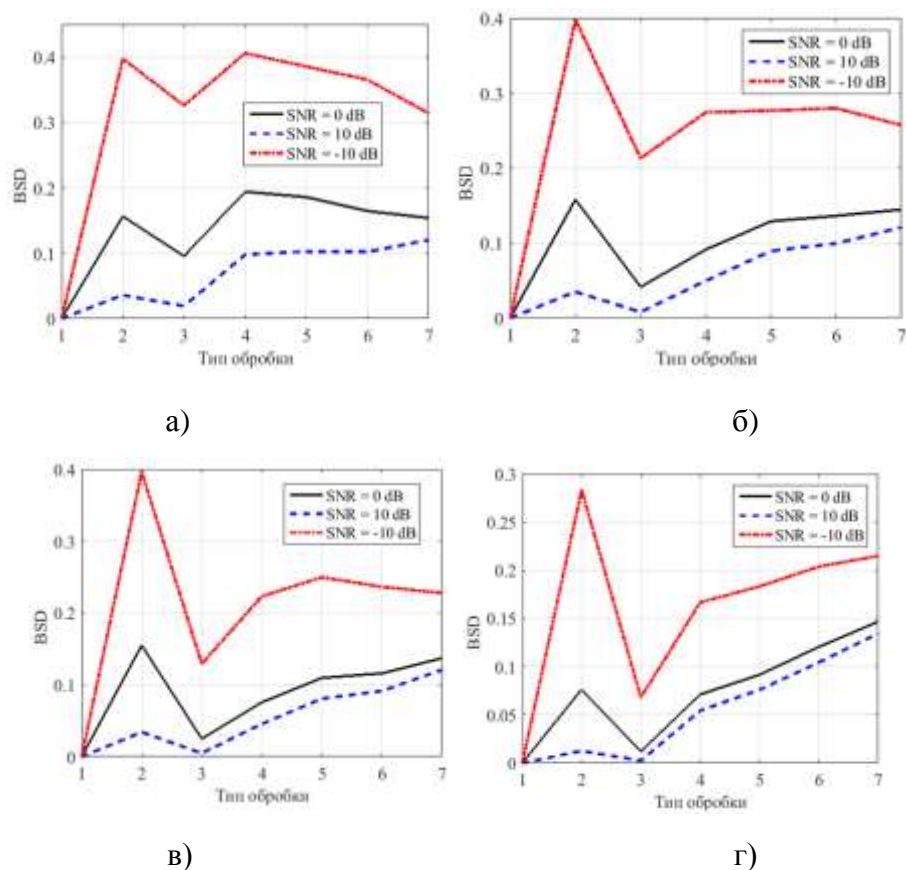


Рис. 4.11. Міра якості BSD: ситуація 1 (а), ситуація 2 (б), ситуація 3 (в), ситуація 4 (г).

Таким чином, в результаті досліджень суб'єктивним та об'єктивним методами можна зробити попередній висновок, що алгоритм DS є більш ефективним, порівняно з алгоритмом LCMV. При цьому, як вже вказувалося, парадоксальним є те, що при використанні алгоритму LCMV якість сигналу виявляється навіть нижчою, ніж для сигналу з виходу одиночного мікрофона.

Досить вірогідним видається припущення, що однією з можливих причин отриманих низькоякісних результатів обробки музичного сигналу можуть бути певні недоліки комп'ютерної програми-прототипу, запозиченої з бібліотеки Matlab.

Для перевірки цього припущення, гармонічний сигнал з частотою 100 Гц та довжиною сигналу 10 секунд було оброблено двома алгоритмами, що досліджуються. Для прискорення часу розрахунку частоту дискретизації встановлено 8 кГц, яка є достатньою для обробки гармоніки з частотою 100 Гц.

Було встановлено такі ж параметри мікрофонного масиву, як при експериментальних дослідженнях. Довжина сегментів, на які поділяється сигнал, виражена у кількості вибірок, становить 1000 відліків за секунду, що відповідає 0.125 секундам. Відгуки DS та LCMV антен наведені на риунках 4.12 та 4.13.

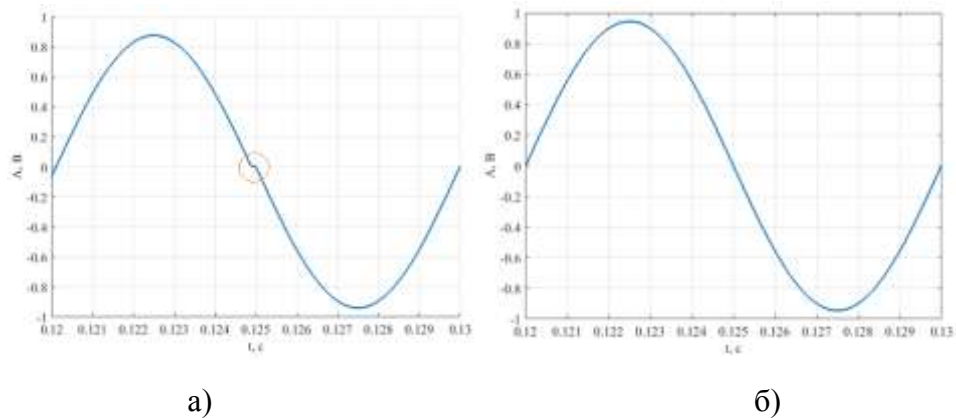


Рис. 4.12. Гармонійний сигнал на виході алгоритму DS: при сегментуванні (а) та без сегментування (б) вхідного сигналу

На рис. 4.12а досить чітко спостерігається спотворення сигналу в місцях склейки оброблених сегментів гармонійного сигналу. Відсутність спотворення сигналу при його обробці без поділення на сегменти продемонстровано на рис. 4.12б.

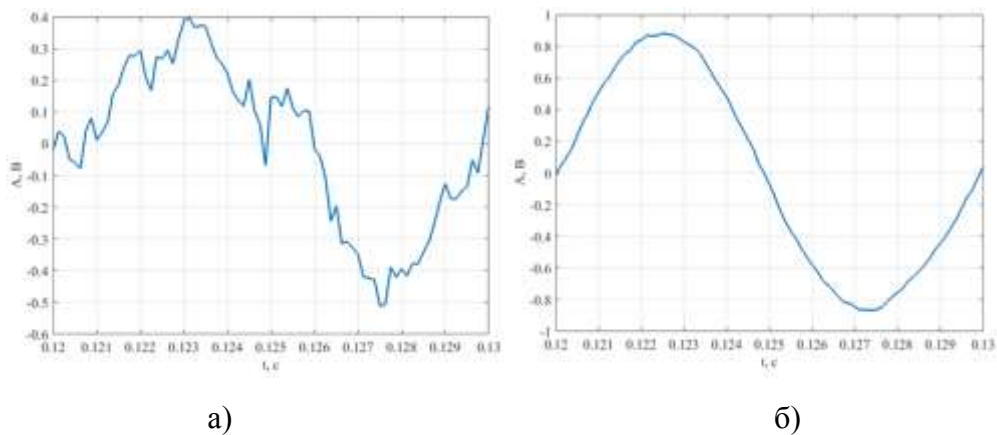


Рис. 4.13. Гармонічний сигнал на виході ФХС LCMV: при сегментуванні (а) та без сегментування (б) вхідного сигналу

Значно вищим, ніж у випадку алгоритму DS, є ступінь спотворення при використанні алгоритму LCMV (рис. 4.13а). На рисунку 4.13б наведено результат обробки гармонічного сигналу за допомогою алгоритму LCMV без поділення на фрейми. Слід, однак, зазначити, що навіть за відсутності поділення на фрейми форма сигналу, отриманого на виході алгоритму LCMV, дещо відрізняється від форми гармонічного сигналу. Можливою причиною цього може бути досить висока складність обчислень при застосуванні алгоритму LCMV, що призводить до помітної похибки обчислень, яка на слух може сприйматися як результат нелінійних спотворень.

Ще одним фактором, що призводить до спотворення сигналу в моменти «склейки» сегментів (як у випадку, що на рис. 4.12а) є некоректність врахування обмеженості смуги пропускання мікрофонів, що входять до складу мікрофонного масиву (рис 4.14).

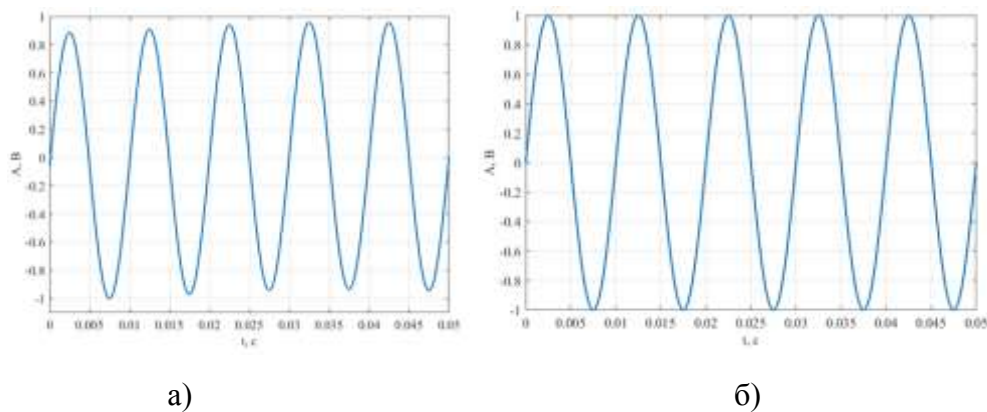


Рис. 4.14. Алгоритм DS, некоректне врахування обмеженості смуги пропускання: обмежена 20 Гц – 20 кГц (а), необмежана (б)

Експериментальна перевірка свідчить, що навіть для алгоритму DS значення показника якості SSNR для смуги пропускання 20 Гц – 20 кГц є пониженим та становить близько 29.3 дБ, у той час як для смуги пропускання 0 Гц – 22050 Гц значення міри SSNR є максимальним та дорівнює 35 дБ, що свідчить про відсутність спотворень сигналу. Тому при подальшому

використанні комп'ютерної програми MicrophoneBeamformerExample.m в якості прототипу слід це врахувати та вважати мікрофони ідеальними, зі смугою пропускання 0-22050 Гц, при частоті дискретизації 44100 Гц.

Що стосується значно сильніших спотворень сигналу на виході алгоритму LCMV, показаних на рис. 4.13а, доречно припустити, що вони спричинені додатковим фактором, а саме, реалізацією алгоритму LCMV в частотній області. Дійсно, якщо вхідний сигнал при обробці ділиться на сегменти, то частина частотних компонентів цих сегментів попадає у вузли сітки частот, що визначаються виразом

$$f = \frac{f_s}{N_{fft}} \cdot i, \quad (4.2)$$

(де значення i – номер вузла сітки частот, f_s – частота дискретизації, N_{fft} – кількість відліків перетворення Фур'є), а частина компонентів – не попадає. Доречно припустити, що «винними» у спотворенні результату є частотні компоненти сегменту, що не попадають у вузли сітки частот.

Для перевірки цього припущення алгоритмом LCMV оброблено два гармонічних сигнали: гармонічний сигнал із частотою, розрахованою за допомогою формули 4.2, яка дорівнює 100.09765625 Гц, та з частотою, що не потрапляє у вузол сітки частот, яка дорівнює 100 Гц. Результати обробки показано на рис. 4.15.

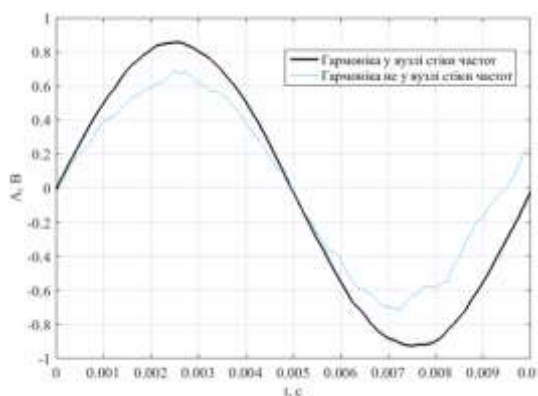


Рис. 4.15. Гармонічний сигнал при обробці алгоритмом LCMV

Як видно з рис. 4.15, сигнал, що не потрапляє у вузол сітки частот перетворення Фур'є, суттєво спотворений за формою та амплітудою.

Таким чином, доходимо висновку, що причинами надмірних спотворень сигналів на виходах DS та LCMV антен, що моделювалися в середовищі Matlab, є недосконалість програмних модулів Matlab, що призводить:

- до некоректного врахування обмеженості АЧХ мікрофонів;
- до суттєвих похибок при посегментній обробці сигналів в частотній області.

4.2.2. Уточнюючі експериментальні дослідження з оцінювання якості музичних сигналів в багатомікрофонних системах

При проведенні уточнюючих експериментальних досліджень параметри мікрофонної решітки було скориговано:

- антена – лінійна, дискретна;
- кількість мікрофонів $N=25$ з кроком 0.02 м;
- кожен мікрофон має АЧХ прямокутної форми з граничними частотами 0 Гц та 22050 Гц;

Крок між мікрофонами обрано з урахуванням значення верхньої граничної частоти музичного сигналу, що є близькою до 8 кГц. (рис. 4.16). Кількість мікрофонів $N=25$ обрано для того, щоб зберегти довжину бази мікрофонної решітки.

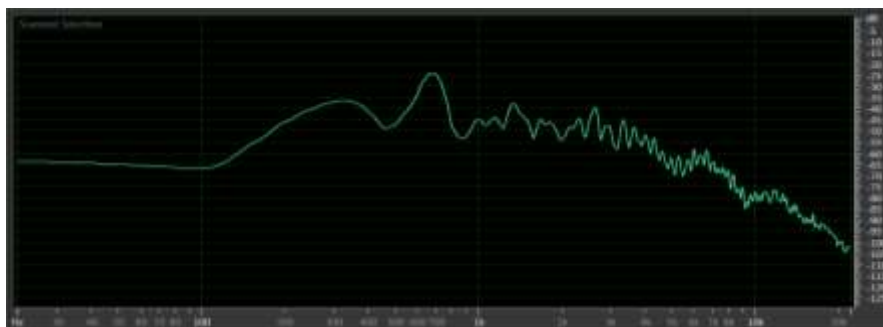


Рис. 4.16. Спектр музичного сигналу (саксофон)

В уточнюючому експерименті розглянуто такі ж самі ситуації джерела завади, як і в попередньому, пілотному експерименті.

Для усунення некоректного обчислення вихідного сигналу за алгоритмами LCMV та DS, спричиненого особливостями програмного забезпечення Matlab, нами запропоновано відмовитися від посегментної обробки сигналів, а також запропоновано вважати, що смуга пропускання становить 0-22050 Гц, а не 20-20000 Гц, як це пропонується в програмному забезпеченні від Matlab.

Усереднені результати суб'єктивного оцінювання в уточнюючому експерименті наведено на рисунках 4.17-4.20. Кількість слухачів становила 25 чоловік, 24 з яких – віком від 20 до 27 років та один 70 років. Всі слухачі – зі здоровим слухом.

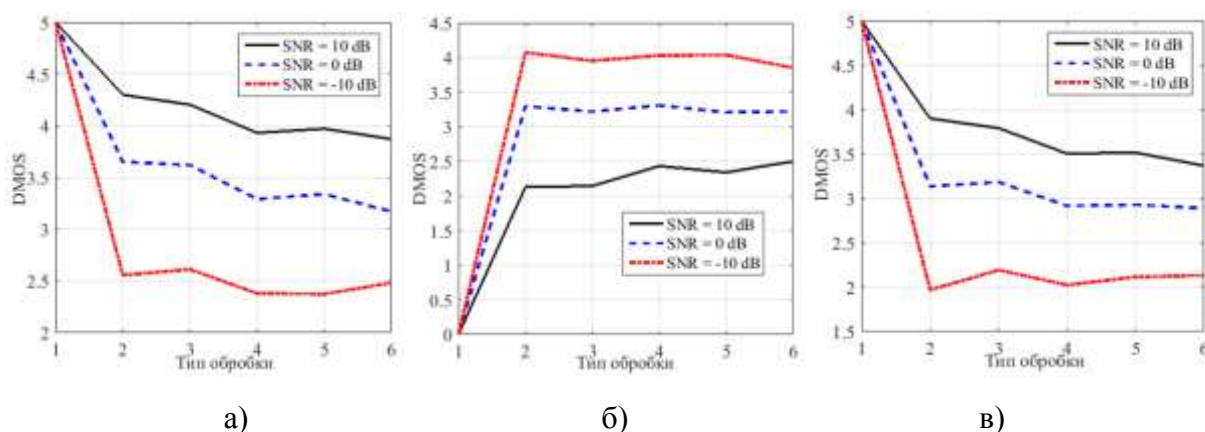


Рис. 4.17. Результати суб'єктивного тесту для ситуації 1 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

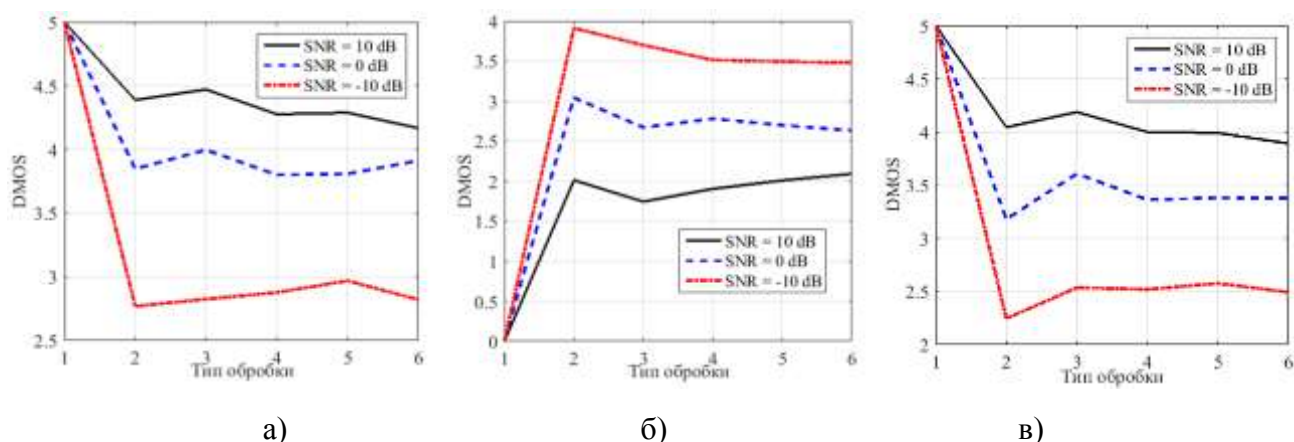


Рис. 4.18. Результати суб'єктивного тесту для ситуації 2 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

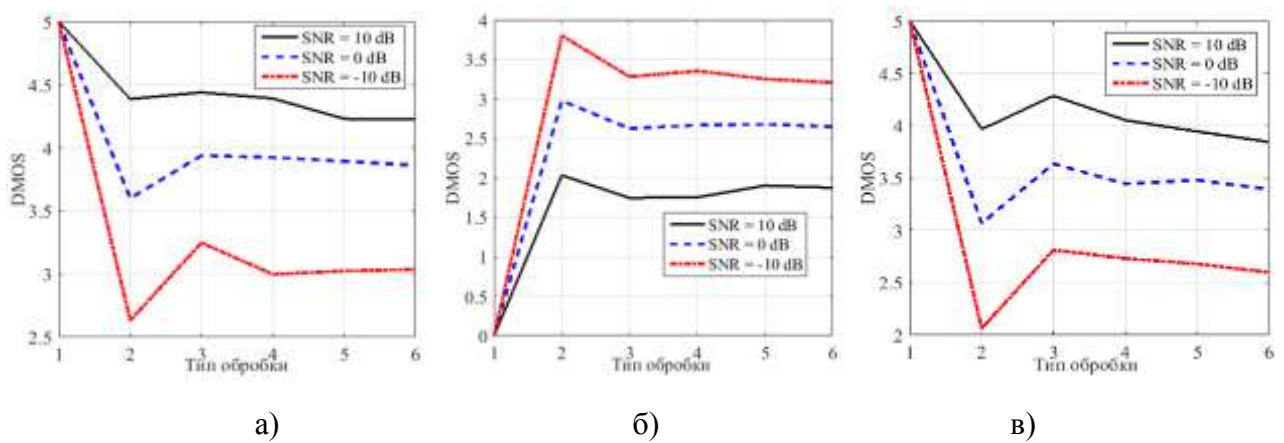


Рис. 4.19. Результати суб'єктивного тесту для ситуації 3 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

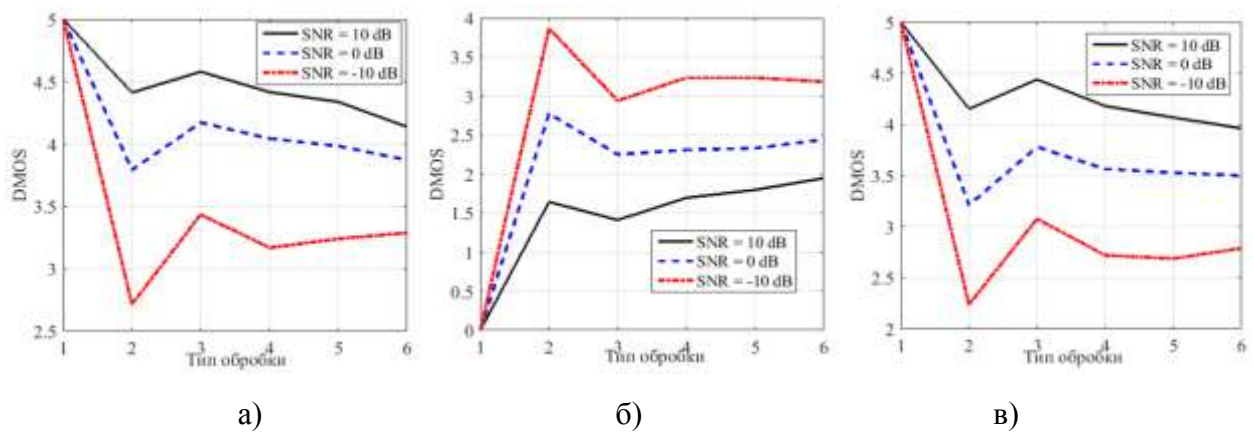


Рис. 4.20. Результати суб'єктивного тесту для ситуації 4 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

Згідно з графіками на рис. 4.17-4.20, суб'єктивні оцінки виявилися досить високими для обох алгоритмів, які досліджувалися, причому відносно незалежними від ситуацій напрямків надходження завади. Така невелика різниця між ситуаціями пояснюється похибками оцінювання. Відносне стандартне відхилення суб'єктивних оцінок представлено на рис. 4.21-4.24. У порівнянні з результатами попереднього експерименту (рис. 4.4-4.7) суттєво зросла оцінка якості сигналу обробленого алгоритмом LCMV, що пояснюється відсутністю значних спотворень музичного сигналу.

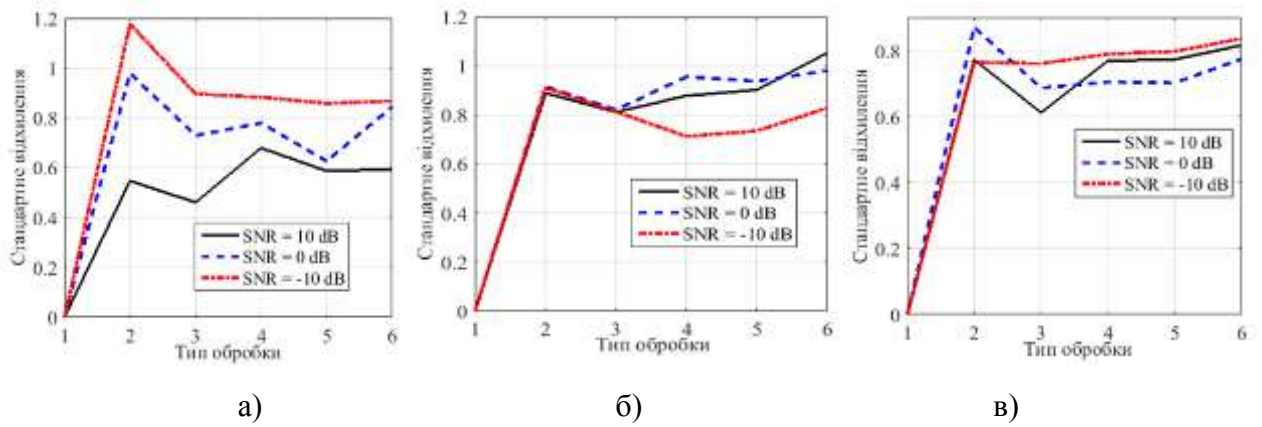


Рис. 4.21. Стандартне відхилення суб'єктивного тесту для ситуації 1 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

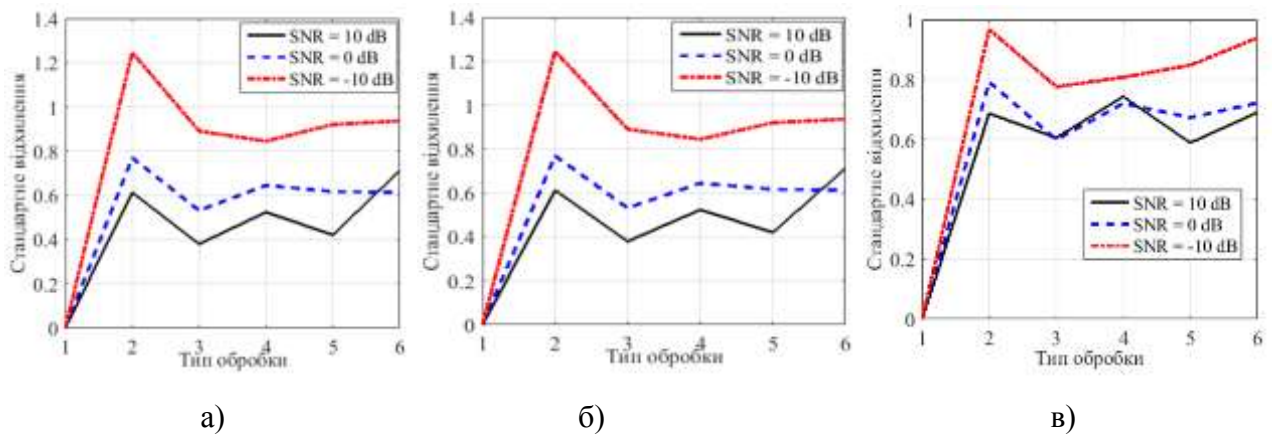


Рис. 4.22. Стандартне відхилення суб'єктивного тесту для ситуації 2 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

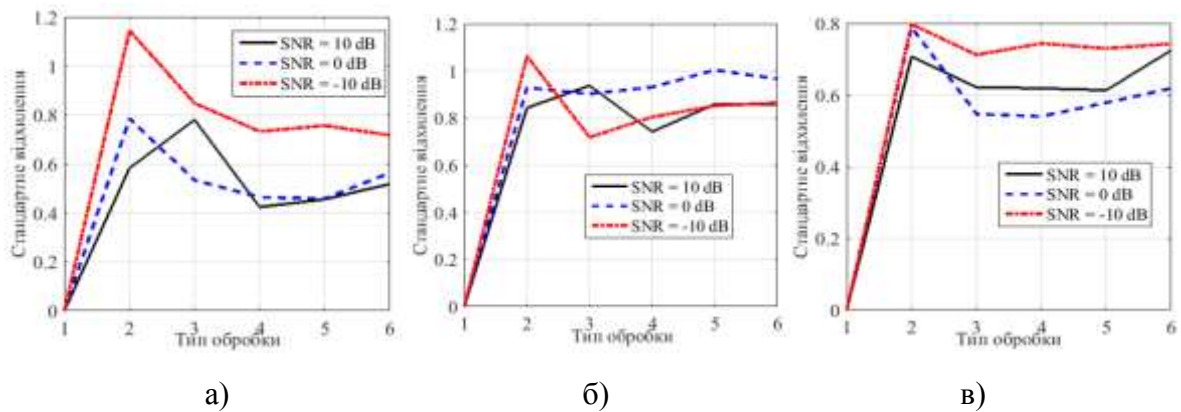


Рис. 4.23. Стандартне відхилення суб'єктивного тесту для ситуації 3 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

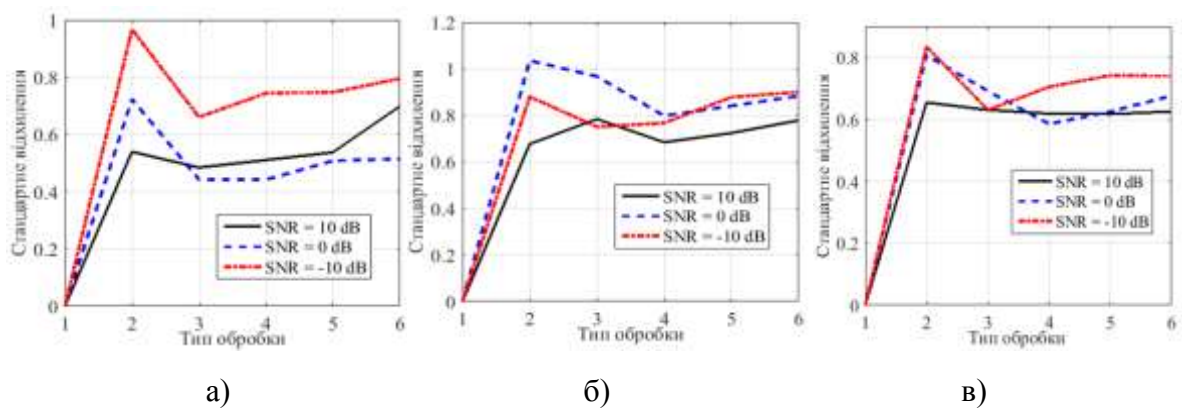
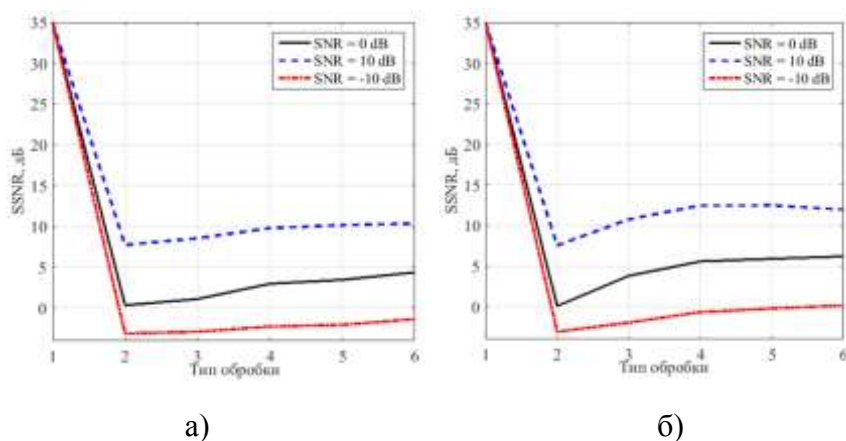


Рис. 4.24. Стандартне відхилення суб'єктивного тесту для ситуації 4 (уточнюючий експеримент): оцінка якості сигналу(а), оцінка рівня завади(б), оцінка загального враження(в).

Достатньо велике значення (від 0.4 до 1.2) відносного стандартного відхилення суб'єктивного оцінювання якості музичного сигналу (рис. 4.21-4.24) можна пояснити недостатньою кількістю слухачів.

Зазначимо, що результати суб'єктивного оцінювання добре узгоджуються з результатами роботи [57], тобто, алгоритм DS забезпечує кращу якість сигналу, порівняно із алгоритмом LCMV.

Результати об'єктивного оцінювання якості сигналів для уточнюючого експерименту представлено на рис. 4.25-4.28.



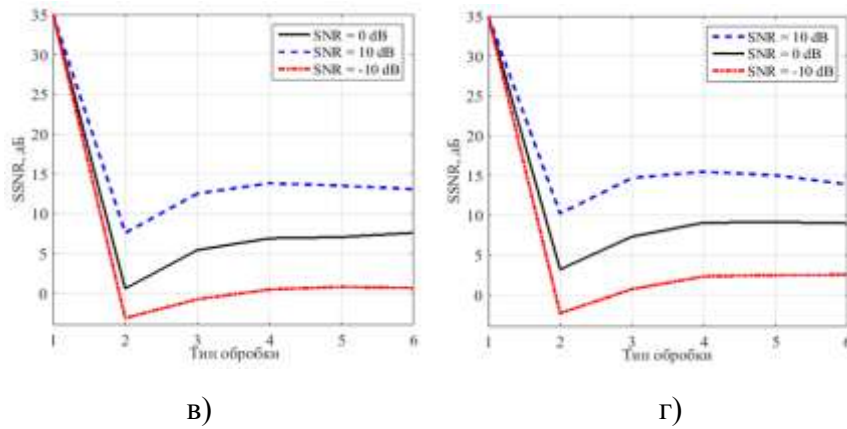
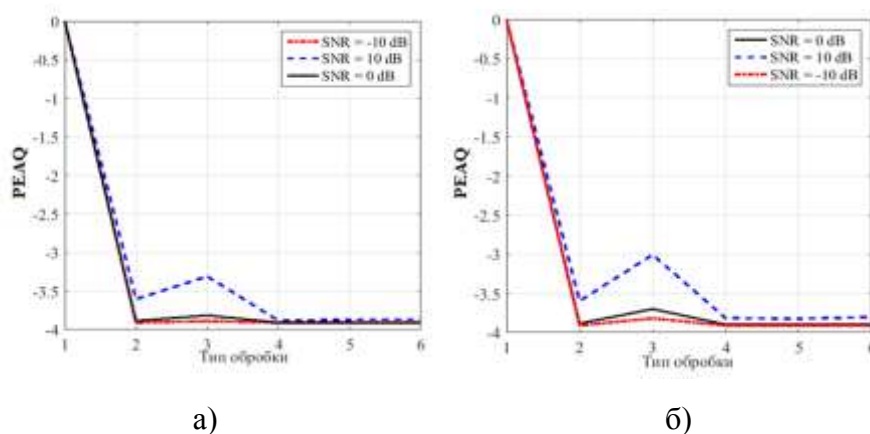


Рис. 4.25. Міра якості SSNR (уточнюючий експеримент): ситуація 1 (а), ситуація 2 (б), ситуація 3 (в), ситуація 4 (г)

Показник міри якості SSNR (рис. 4.25), для алгоритму LCMV, збільшився в 2.5 рази, у порівнянні з попереднім експериментом, що пояснюється відсутністю вище обговорюваних недоліків в конфігурації параметрів даного експерименту. Порівнюючи графіки на рис. 4.25 та рис. 4.8, коригування параметрів мікрофонних решіток на міру якості SSNR для алгоритму DS не вплинуло, що пояснюється тим, що система цього алгоритму є лінійною, а отже нелінійні спотворення відсутні в обох експериментах.

Значно цікавішим є те, що, згідно з мірою SSNR, якість сигналів для алгоритму LCMV виявилася вищою за таку для алгоритму DS. Цей результат суперечить як результатам суб'єктивного оцінювання, так і результатам роботи [57].



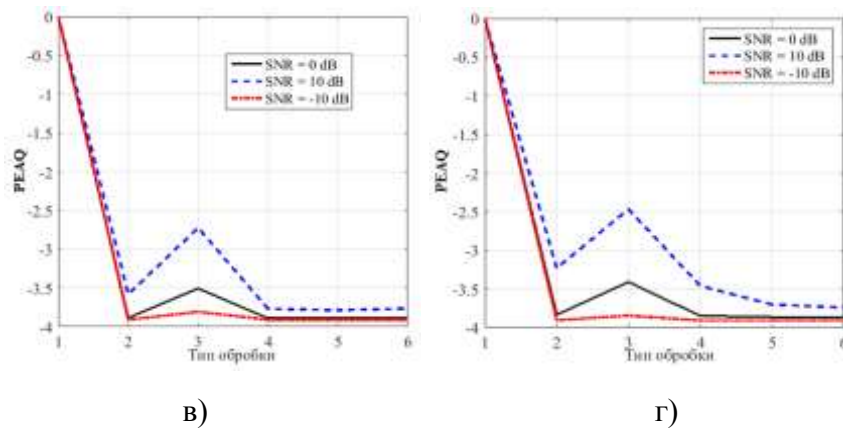
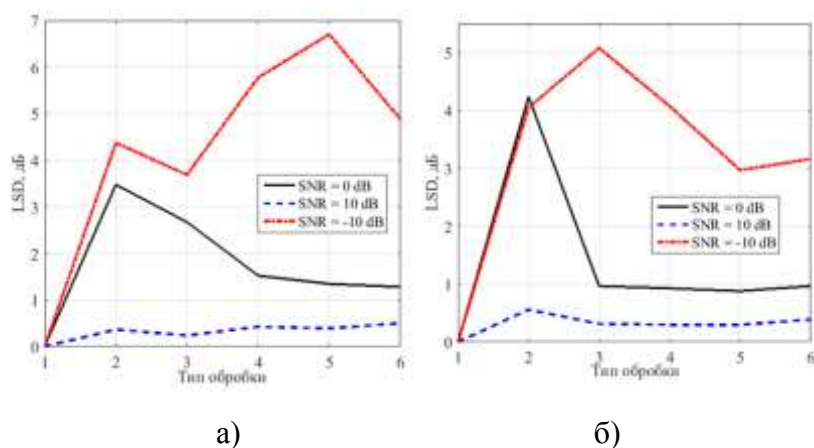


Рис. 4.26. Міра якості PEAQ (уточнюючий експеримент): ситуація 1 (а), ситуація 2 (б), ситуація 3 (в), ситуація 4 (г)

Згідно графікам (рис. 4.26) міра якості PEAQ, порівняно з графіками на рис. 4.9, зросла на 1-цю тільки для алгоритму DS. Однак, показник PEAQ є індикатором поліпшення роботи алгоритму DS для ситуацій, коли співвідношення сигнал/шум становить 0 дБ, що пояснюється зменшенням нелінійних спотворень. Загалом низька оцінка міри PEAQ пояснюється внесенням спектральних спотворень (рис. 4.27) в музичний сигнал – тобто наявністю шумової складової, яка залишилася в певному співвідношенні (див. рис. 4.25) після обробки масивом мікрофонів.

Значно цікавішим є те, що, згідно з мірою PEAQ, якість сигналів для алгоритму LCMV виявилася нижчою за таку для алгоритму DS. Цей результат узгоджується як із результатами суб'єктивного оцінювання, так і з результатами роботи [57].



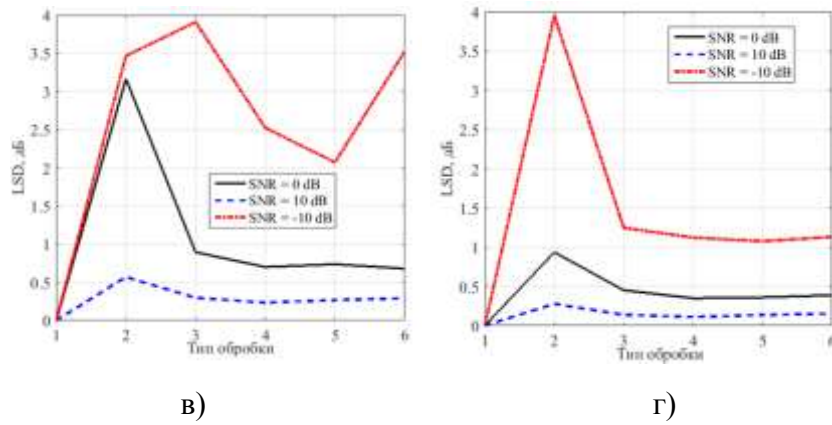
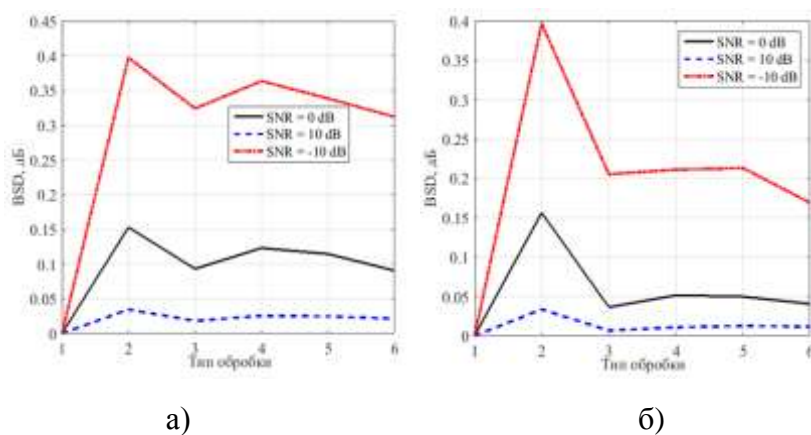


Рис. 4.27. Міра якості LSD (уточнюючий експеримент): ситуація 1 (а), ситуація 2 (б), ситуація 3 (в), ситуація 4 (г)

Найбільші значення показника LSD спостерігалися при значенні $\text{SNR} = -10$ дБ (рис. 4.27); причому несподівано, що для ситуацій 2 та 3 значення міри LSD є найбільшим для алгоритму DS. У порівнянні з попереднім експериментом (рис. 4.10), очікуваним виявилось значне зменшення показника LSD для алгоритму LCMV майже до значення такого алгоритма DS, в ситуаціях 2, 3 та 4, а у випадку ситуації 1 – взагалі менше на 1 дБ, що пояснюється більш гострою та відносно незалежною від частоти ХС алгоритма LCMV.

Значно цікавішим є те, що, згідно міри LSD, якість сигналів для алгоритму LCMV є в одних ситуаціях вищою за таку для алгоритму DS, а в інших ситуаціях – нижчою. Цей результат лише частково узгоджується як із результатами суб'єктивного оцінювання, так і з результатами роботи [57].



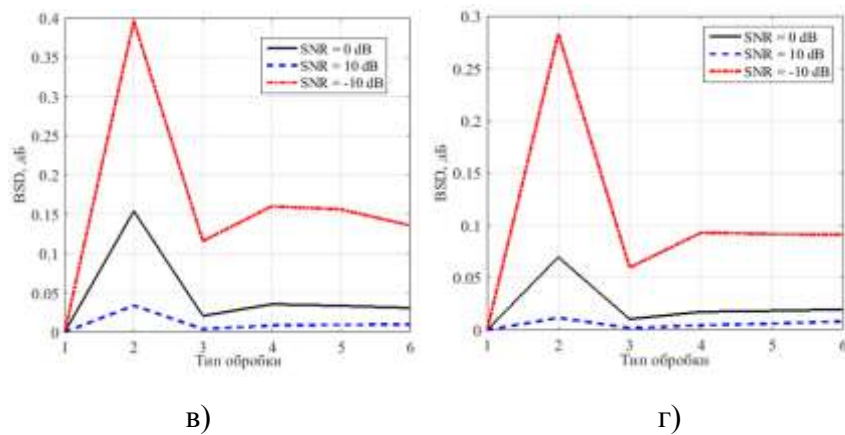


Рис. 4.28. Міра якості BSD (уточнюючий експеримент): ситуація 1 (а), ситуація 2 (б), ситуація 3 (в), ситуація 4 (г)

Поведінка міри BSD виявилась досить передбачуваною. Так само, як і у випадку з показником міри LSD (рис. 4.27), коригування параметрів не вплинуло на значення міри BSD для алгоритму DS за відсутності нелінійних спотворень.

Значно цікавішим є те, що, згідно з мірою BSD, якість сигналів для алгоритму LCMV виявилася нижчою за таку для алгоритму DS. Цей результат узгоджується як із результатами суб'єктивного оцінювання, так і з результатами роботи [57].

Таким чином, одержані результати суб'єктивного та об'єктивного оцінювання якості сигналів для алгоритмів DS та LCMV в цілому узгоджуються з висновками роботи [57] щодо переваги алгоритму DS над алгоритмом LCMV. Разом із тим, дослідження показали, що при застосуванні об'єктивної міри SSNR такого узгодження немає, а при застосуванні міри LSD узгодження є частковим. Отримані результати, з одного боку, свідчать про «проблемний характер» таких простих для обчислення мір як SSNR та LSD, про що вже говорилося в розділі 2. З іншого боку, отримані результати підтверджують висновки попередніх дослідників щодо програшу, за критерієм якості сигналів, складного алгоритму LCMV більш простому алгоритму DS. Цей результат є дуже цікавим з точки зору теорії цифрової просторово-часової обробки сигналів, а також є дуже корисним з точки зору практичного використання алгоритмів просторово-часової обробки музичних сигналів.

Висновки до розділу 4

1. Виконано експериментальне порівняння, за критерієм якості сигналів, алгоритмів LCMV та DS просторово-часової обробки музичних сигналів, що дозволяє зробити висновки щодо доцільності використання складних алгоритмів LCMV замість простіших алгоритмів DS.
2. Вперше отримані результати суб'єктивного оцінювання якості музичних сигналів для алгоритмів LCMV та DS свідчать, що більш складний алгоритм LCMV поступається простішому алгоритму DS за якістю сформованих сигналів. Показано, що при застосуванні обох алгоритмів найбільш проблемним є питання вибору розміру сегментів при посегментній обробці сигналів, а також показано, що алгоритм LCMV є набагато чутливішим до такого вибору.
3. Результати об'єктивного оцінювання якості музичних сигналів для алгоритмів LCMV та DS свідчать, що об'єктивні міри BSD та PEAQ адекватно відображають якість сигналів, сформованих алгоритмами LCMV та DS, оскільки цілком узгоджуються з результатами суб'єктивного оцінювання та результатами попередніх дослідників. Разом із тим показано, що питання застосування об'єктивних мір SSNR та LSD є проблематичним і тому має бути досліджене додатково.

ВИСНОВКИ

Основні результати дисертаційної роботи полягають у наступному.

1. Вперше для реальних акустичних сигналів отримано кількісні суб'єктивні та об'єктивні оцінки ступеню впливу нелінійності фазової характеристики тракту на якість мовних та музичних сигналів, що дозволяє сформулювати адекватні вимоги до ступеня спотворення акустичних сигналів, а саме:

- оскільки фазові спотворення музичних сигналів значно менш помітні на слух, ніж такі для мовних сигналів, для якісного відтворення музичних сигналів різниця часу групової затримки на низьких та високих частотах не має перевищувати 80 мс, а мовних сигналів – 40 мс;

- слухова система людини є чутливою до знаку різниці часу групової затримки для низьких та високих частот, що має враховуватися при налагодженні систем відтворення мовних та музичних сигналів.

2. Вперше запропоновано використання кумулянтного коефіцієнту статистичного розподілу миттєвих значень акустичного сигналу у вигляді коефіцієнта ексцесу та його функціональних перетворень як об'єктивних мір ступеня кліпування та якості мовних та музичних сигналів, спотворених кліпуванням. Дана пропозиція дозволяє суттєво (в 200-300 разів) знизити вимоги до розмірності вектору параметрів, що оцінюється при виявленні кліпування і таким чином значно скоротити час відповідних обчислень без зниження достовірності результатів.

3. Вперше вказано на необхідність врахування підвищеної чутливості простої в обчислювальному плані об'єктивної міри якості сигналів у вигляді лог-спектральних спотворень до спектрального складу акустичних сигналів, що аналізуються. Отримані рекомендації з урахування такої підвищеної чутливості дозволяють підвищити достовірність висновків при оцінюванні якості спотворених мовних та музичних сигналів.

4. Уточнено висновки про принципову можливість підвищення достовірності об'єктивного оцінювання якості спотворених акустичних сигналів з використанням такої міри як сегментне відношення сигнал-шум та показано, що такий недолік як чутливість сегментного відношення сигнал-шум до похибки вирівнювання сигналів може бути нейтралізовано збільшенням частоти дискретизації шляхом інтерполяції щонайменше в 5 разів для музичних сигналів, тоді як для мовних сигналів частоту дискретизації достатньо збільшувати в 2 рази.

5. Удосконалено технологію автоматизації суб'єктивного оцінювання розбірливості мови, спотвореної шумом та реверберацією, артикуляційним методом, для чого запропоновано при фіксації почутих односкладових слів записувати повне слово замість дописування відсутніх літер, а також розроблено та протестовано відповідне програмне забезпечення. Запропонована технологія є значно простішою у використанні, порівняно із зарубіжними аналогами, що дозволяє значно пришвидшити та знизити вартість проведення артикуляційних досліджень без зниження їх достовірності.

6. Уточнено межі застосування певного переліку об'єктивних оцінок якості мови замість суб'єктивних оцінок розбірливості зашумленої мови, що дозволяє при значеннях відношення сигнал-шум, більших за 0 дБ, оцінювати розбірливість мови через оцінювання якості мови й таким чином значно пришвидшити процедуру об'єктивного оцінювання розбірливості мови без зниження достовірності оцінок.

7. Уточнено попередні висновки про значну роль ранніх відбиттів звуку в приміщеннях, для чого виконано порівняння результатів артикуляційних досліджень при монауральному та бінауральному режимах прослуховування мовних сигналів, спотворених шумом та реверберацією. Отримані результати свідчать, що ранні відбиття хоча і є корисними з енергетичної точки зору, оскільки призводять до збільшення відношення сигнал-шум на 6-9 дБ, проте вони є шкідливими з інформаційної точки зору, оскільки розбірливість мови

знижується на 3-5% на відстанях 5-9 м до джерела звуку, що спричинене нерівномірністю частотної характеристики приміщення.

8. Уточнено результати зіставлення, шляхом суб'єктивного та об'єктивного оцінювання якості звуку для кількох алгоритмів обробки спотворених шумом музичних сигналів в багатомікрофонних масивах та показано, що при використанні простішого, з обчислювальної точки зору, алгоритму "сум та затримок" можна отримати більш якісні сигнали, ніж при використанні складних алгоритмів, що забезпечують незалежні від частоти діаграми підвищеної спрямованості.

9. Проведено попередній аналіз причин погіршеної якості сигналів для алгоритмів, що забезпечують незалежні від частоти діаграми підвищеної спрямованості, та вказано, що таке погіршення спричинене значною похибкою обчислень при реалізації цих алгоритмів, причому ця похибка зростає при зменшенні довжини сегментів сигналу та при зменшенні смуги пропускання мікрофонів відносно частоти дискретизації. Отримані результати мають сприяти подальшим дослідженням та застосуванню багатомікрофонних систем просторово-часової обробки звукових сигналів, спотворених шумом та реверберацією.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Perceptual Evaluation of Speech Quality (PESQ) ITU-T Recommendations P.862, P.862.1, P.862.2. Version 2.0 - October 2005.
- [2] D. Câmpeanu and A. Câmpeanu, «PEAQ – an objective method to assess the perceptual quality of audio compressed files», URL: <https://www.researchgate.net/publication/228944316>.
- [3] N. Cote, «Integral and diagnostic intrusive prediction of speech». - Springer-Verlag Berlin Heidelberg. – 2011. – 267 p.
- [4] J.S. Bradley, «The evolution of newer auditorium acoustics measure»/ Canadian Acoustics / Acoustique Canadienne 18(4) P. 13-23, 1990.
- [5] A. Wilson, B. Fazenda, «Perception & evaluation of audio quality in music production»/Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13), Maynooth, Ireland, September 2-5, 2013.
- [6] А. Продеус, В. Дидковский, М. Дидковская, «Акустическая экспертиза и коррекция коммуникационных каналов». Монография. LAP LAMBERT Academic Publishing, OmniScriptum GmbH & Co. KG, Saarbrücken, Deutschland, 2017, ISBN: 978-3-330-04591-0
- [7] К.С. Замша, Б.В. Лозинский, Ю.А. Митяй, Е.С. Степановская, А.Н. Продеус. «Объективное и субъективное оценивание качества речевых сигналов с ограниченной полосой частот». - Electronics and Communications, Vol. 21, № 1(90), 2016. – P. 18-26
- [8] А.Н. Продеус, К.П. Пилипенко, А.Я. Калужный, С.Г. Бартенев, «Оценка влияния нелинейности фазовой частотной характеристики системы на качество речевых сигналов», *Электроника и связь*, т.20, №2(85), 2015. - С.33-40.
- [9] А.Н. Продеус, Н.В. Богданова, «Оценка влияния нелинейности фазовой частотной характеристики тракта на качество музыкальных сигналов», *Electronics and Communications*, Vol. 20, No. 4(87), 2015. – P. 29-35.

- [10] C. J. Brian Moore. «Effects of Bandwidth, Compression Speed, and Gain at High Frequencies on Preferences for Amplified Music», Trends Amplif. 2012 Sep; 16(3): 159–172 p.
- [11] C. Raffel, P. W. Daniel Ellis, «Estimating timing and channel distortion across related signals», 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)
- [12] T. Otani, «Clipping detection device and method»/ Otani T., Tanaka M., Ota Y., Ito S. // Patent US 8,392,199 B2, Int.Cl. G10 19/00, 2013.
- [13] Г.Р. Аванесян, «Метод и устройство для оценки и индикации искажений выходного сигнала усилителя звуковой частоты (индикация перегрузки)» / Аванесян Г.Р. // Патент RU 2274868 C2, МПК G01R 23/20, G01R 19/165, 2006.
- [14] X. Liu, J. Jia, L. Cai, «SNR estimation for clipped audio based on amplitude distribution»/ Liu X., Jia J., Cai L. // Proc. of the IEEE 9th Int. Conf. on Natural Computation (ICNC), 23-25 July, 2013. - IEEE, 2013. – DOI: 10.1109/ICNC.2013.6818205.
- [15] С.В. Алейник, «Метод оценки уровня клиппирования речевых сигналов»/ Алейник С.В., Матвеев Ю.Н., Раев А.Н. // Научно-технический вестник информационных технологий, механики и оптики. - 2012. - № 3 (79). - С. 79–83.
- [16] S.V. Aleinik, «Detection of clipped fragments in acoustic signals» / Aleinik S.V., Matveev Yu.N., Sholokhov A.V. // International Journal of Computer and Information Engineering . – 2014. – Vol. 8, No. 2. - P. 286-292.
- [17] F. Bie, «Detection and reconstruction of clipped speech for speaker recognition» / Bie F., Wang D., Wang J, Zheng T.F. // Speech Communication. - September 2015. – Vol. 72. – P. 218-231. – DOI: 10.1016/j.specom.2015.06.008.
- [18] C. Laguna, «An efficient algorithm for clipping detection and declipping audio» / Laguna C., Lerch A. // AES 141st Convention. - September 29-October 2, 2016. - Los Angeles, USA.

- [19] A. Poorjam, «Dominant Distortion Classification for Pre-Processing of Vowels in Remote Biomedical Voice Analysis»/ Poorjam A., J. Jensen, M. Little, Christensen M. // INTERSPEECH 2017. - August 20–24, 2017. - Stockholm, Sweden. - DOI : 10.21437/Interspeech.2017-378
- [20] M. Kendall, «The Advanced Theory of Statistics: Distribution theory» / Kendall M., Stuart A. - London: Wiley,1977.
- [21] J. Moors, «The Meaning of Kurtosis»/Darlington Reexamined / Moors J. // The American Statistician. – 1986. - 40:4. – P. 283-284.
- [22] V. Arora, «Probability distribution estimation of music signals in time and frequency» / Arora V., Kumar R. // Proc. of the IEEE 19th Int. Conf. on Digital Signal Processing (DSP-2014), 20-23 August 2014, Hong Kong. - IEEE, 2014. – P. 409-414. DOI: 10.1109/ICDSP.2014.6900696.
- [23] GOST 16600-72. The transmission of speech through the radiotelephone communications. Requirements to intelligibility of speech and methods of articulatory measurements, Standartinform, Moscow, 2007, 77 p. (in Russian).
- [24] GOST R 50840-95. Transmission of speech through communication paths. Methods for assessing quality, intelligibility and recognizability, Gosstandart of Russia, Moscow, 1996, 234 p. (in Russian).
- [25] V.G. Mikhailov, "Diagnostic articulation tables," Translated from Akusticheskii Zhurnal, Vol. 48, No. 5, 2002, pp. 705–712 [Online]. Available: <https://link.springer.com/article/10.1134/1.1507211>.
- [26] E. Arkhipova, V. Zhuravlev, V. Kumeiko, “Articulation tables of words of the Ukrainian language,” Legal, normative and metrological provision of the information security system in Ukraine, issue 2(19), 2009, pp.13-17 (in Russian).
- [27] A. Arkhipov, E. Arkhipova, “Analysis and processing of articulation tests data,” Protection of Information, No. 4, 2012, pp. 34-42 (in Russian).
- [28] ANG-2200. Acoustic Noise Generator. Owner’s Guide. Research Electronics, Intl. 455 Security Place Algood, TN 38506, USA. 33 p. [Online]. Available: https://reiusa.net/wp-content/uploads/2017/11/ANG_Manual_revG.pdf

- [29] G. Li, «Relationship between speech recognition in noise and sparseness» / Li G., Lutman M., Wang S., Bleek S. // International Journal of Audiology. – 2012. – No. 51. – P. 75–82.
- [30] N. Cote, «Integral and diagnostic intrusive prediction of speech»/Cote N. - Springer-Verlag: Berlin-Heidelberg, 2011.
- [31] J. Beerends, E. Larsen, N. Iyer, and J. Vugt, «Measurement of speech intelligibility based on the PESQ approach», Proceedings of the Workshop Measurement of Speech and Audio Quality in Networks (MESAQIN), Prague, Czech Republic, June 2004.
- [32] J. Beerends, S. Wijngaarden, and R. Buuren, «Extension of ITU-T recommendation P.862 PESQ towards measuring speech intelligibility with vocoders», In New Directions for Improving Audio Effectiveness, Proceedings of the RTO-MP-HFM-123, Neuilly-sur-Seine, France, 2005, pp. 10-1–10-6.
- [33] J. Ma, Y. Hu, and P. Loizou, «Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions», J. Acoust. Soc. Am., Vol. 125, No. 5, May 2009, pp. 3387-3405.
- [34] A. Prodeus, V. Didkovskyi, M. Didkovska. Acoustic examination and correction of communication channels: Monograph, LAP LAMBERT Academic Publishing, Deutschland, 2017, 364 p. ISBN-10: 3330045914, ISBN-13: 978-3330045910 (Russian Edition).
- [35] Yu. Makarov, A. Khorev, "To the evaluation of the effectiveness of protection of acoustic (speech) information," Special Technique, No. 2, 2000 [Online]. Available:
http://www.ess.ru/sites/default/files/files/articles/2000/05/2000_05_05.pdf
(in Russian).
- [36] Bradley, J. S.; Sato, H.; Picard, M. On the importance of early reflections for speech in rooms. Journal of Acoustical Society of America, v. 113, no. 6, June 2003, pp. 3233-3244.

- [37] I. Arweiler, J. Buchholz, T. Dau, Speech intelligibility enhancement by early reflections. - Proceedings of 2nd International Symposium on Auditory and Audiological Research (ISAAR 2009), August 2009, Elsinore, Denmark.
- [38] Yang, W. and Bradley, J. (2009), "Effects of room acoustics on the intelligibility of speech in classrooms for young children," J. Acoust. Soc. Am. 125 (2), P. 922–933.
- [39] Sergey Naida. Acoustic Theory Problems Of Speech Production In The Light Of The Discovery Of The Formula For The Middle Ear Norm Parameter. Proc. of IEEE 35th Int. Sc. Conf. Electronics and Nanotechnology (ELNANO), pp. 347-350, 21-24 April 2015, Kyiv, Ukraine. DOI: 10.1109/ELNANO.2015.7146907.
- [40] Sergey Naida. New Possibilities of the Tympanoplasty After the Discovery of the Formula for the Middle Ear Norm Parameter. Proc. of IEEE XI-th Int. Conf. Perspective technologies and methods in MEMS design (MEMSTECH 2015), pp.96-101, 2-6 September 2015, Lviv - Polyana, Ukraine.
- [41] W. Ahnert and W. Schmidt, Fundamentals to perform acoustical measurements. Appendix to EASERA. Berlin, 2005.
- [42] J. Blauert Ed., The technology of binaural listening. Springer, Berlin–Heidelberg–New York, 2013.
- [43] B. W. Darren, A. Rodney, R. Kennedy, C. Williamson, «Constant Directivity Beamforming»/Springer-Verlag Berlin Heidelberg, 2001. – P. 21-35.
- [44] J. Bitzer, K. Uwe Simmer, Superdirective Microphone Arrays (in M. Brandstein, D. Ward (Eds.), Microphone Arrays. Signal Processing Techniques and Applications). Springer-Verlag, Berlin, Heidelberg, 2001.
- [45] K. Uwe Simmer, J. Bitzer, C. Marro, «Post-Filtering Techniques»/Springer-Verlag Berlin Heidelberg, 2001. – P. 56-77.
- [46] D. N. Zotkin, R. Duraiswami, Nail A. Gumerov, «Sound field decomposition using spherical microphone arrays»/Perceptual Interfaces and Reality Lab, UMIACS, University of Maryland, College Park, ICASSP 2008

- [47] A. Farina, A. Capra, L. Chiesi, L. Scopece, «A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in post production»/AES 40th International Conference, Tokyo, Japan, 2010
- [48] N. Peters, A. W. Schmeder, «Beamforming using a spherical microphone array based on legacy microphone characteristics»/ International Conference on Spatial, Audio Detmold, Germany, 2011.
- [49] Y. Sasaki, M. Kabasawa, S. Thompson, S. Kagami, K. Oro, «Spherical Microphone Array for Spatial Sound Localization for a Mobile Robot»/ Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference.
- [50] A. Avni¹, J. Ahrens, M. Geier, S. Spors, H. Wierstor and B. Rafaely, «Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution»/ The Journal of the Acoustical Society of America , 2013
- [51] A. Jorge, T. Lopez, K. Wakayama, Sh. Sakamoto, Yo-iti Suzuki Hideaki Takada and M. Okamoto, «Recording of extended sound fields using spherical microphone arrays and a-priori knowledge of the sound source positions»/ J. Acoust. Soc. Am. 139, 2108 (2016)
- [52] H. Hacıhabiboğlu, Z. Cvetkovirc, «Panoramic recording and reproduction of multichannel audio using a circular microphone array»/in Proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA '09), pp. 117–120, NewPaltz,NY, USA, October 2009.
- [53] D. Pavlidi, M. Puigt, A. Griffin, A. Mouchtaris, «Real-time multiple sound source localization using a circular microphone array based on single-source confidence measures»/in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '12), pp. 2625–2628,March 2012.
- [54] Falk-Martin Hoffmann, Filippo Maria Fazi, «Circular microphone array with tangential pressure gradient sensors»/IEEE, 2014
- [55] A. Alexandridis, A. Griffin, A. Mouchtaris, «Capturing and Reproducing Spatial Audio Based on a Circular Microphone Array»/ Journal of Electrical and

- [56] W. Zhang, P.N. Samarasinghe, H. Chen, T.D. Abhayapala. «Surround by Sound: A Review of Spatial Audio Recording and Reproduction», Applied Sciences, No. 7, 532, 2017, pp. 1-19.
- [57] P. Coleman, P.J. B. Jackson, and J. Francombe, «Audio object separation using microphone array beamforming», 138th Convention, May 7, Warsaw, Poland, 2015.
- [58] O. Schreer, G. Thomas, O.A. Niamut, J-F. Macq, A. Kochale, J-M. Batke, J. Ruiz Hidalgo, R. Oldfield, B. Shirley, G. Thallinger, «Format-agnostic Approach for Production, Delivery and Rendering of Immersive Media», Format-agnostic Approach for Production, Delivery and Rendering of Immersive Media“, NEM Summit 2011, Torino, Italy, 27th September, 2011.
- [59] A. Farina, A. Capra, L. Chiesi, L. Scopece, «A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production», AES Conference:40th International Conference: Spatial Audio: Sense the Sound of Space, Paper Number 3-1, October 2010.
- [60] A. Farina, A. Amendola, L. Chiesi, A. Capra, S. Campanini, «Spatial PCM Sampling: A New Method for Sound Recording and Playback», 52nd AES Conference: International Conference: Sound Field Control - Engineering and Perception, Paper Number 7-2, September 2013.
- [61] A. Farina, S. Campanini, L. Chiesi, A. Amendola, L. Ebri, «Spatial Sound Recording with Dense Microphone Arrays», AES 55th International Conference, Helsinki, Finland, 2014 August 27–29, pp. 1-8.
- [62] O. Kirkeby, P. A. Nelson, «Digital Filter Design for Inversion Problems in Sound Reproduction», AES, vol. 47, no. 7/8 (1999 July/August).
- [63] O. Kirkeby, P. Rubak, A. Farina, «Analysis of ill-conditioning of multi-channel deconvolution problems», 106th AES Convention, Munich, Germany - May 8-11, 1999.

- [64] P. Kabal, «An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality». TSP Lab Technical Report, Dept. ECE, McGill Univ. May, 2002
- [65] D.B. Ward, R.A. Kennedy, and R.C. Williamson, «Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns», J. Acoust. Soc. Amer., vol. 97, no. 2, pp. 1023-1034, Feb. 1995.
- [66] И. В. Котвицкий, А. Н. Продеус, «Объективное и субъективное оценивание качества речевых и музыкальных сигналов, подвергнутых фазовым искажениям». - ISSN 1811-4512. *Электроника и связь*. 2016, № 2. С. 25-31.
- [67] A.M.Prodeus, I.V. Kotvytskyi, A.A.Ditiashov, «Assessment of clipped speech quality», Electronics and Control Systems, 2018, N 4(58), P. 11-18. DOI:10.18372/1990-5548.58.13504.
- [68] А. Продеус, А. Вітик, О. Дворник, І. Котвицький, О. Чайка, М. Ярошенко, «Суб'єктивне оцінювання розбірливості мови на тлі шуму та реверберації». Мікросистеми, електроніка та акустика. Том 23, № 2, 2018. ISSN 2523-4447, DOI: 10.20535/2523-4455.2018.23.2.128820. С. 66-73.
- [69] М.П. Юраков, І.В. Котвицький, А.М. Продеус, «Оцінювання якості музичних сигналів, обмежених за смугою частот». ISSN 2523-4447. Мікросистеми, Електроніка та Акустика, 2018, vol. 23, № 4,. DOI: 10.20535/2523-4455.2018.23.4.133785. С. 58-64.
- [70] A. Prodeus, K. Bukhta, P. Morozko, O. Serhiienko, I. Kotvytskyi, O. Dvornyk, «Automated Subjective Assessment of Speech Intelligibility in Various Listening Modes», ISSN 2523-4447. Microsystems, Electronics and Acoustics, 2018, vol. 23, no. 3, P.49-57.
- [71] В.С. Дидковский, М.В. Дидковская, А.Н. Продеус, «Акустическая экспертиза каналов речевой коммуникации». Монография. Киев: Имэкс-ЛТД, 2008.
- [72] A. Prodeus, “Reducing Sensitivity of Segmental Signal-to-Noise Ratio Estimator to Time-Alignment Error,” International Journal of Electrical and Electronic Science. August 20, 2015. Vol. 2(2). P. 31-36.

- [73] M. Souden, J. Benesty, and S. Affes, "A study of the lcmv and mvdr noise reduction filters," *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4925–4935, 2010.
- [74] A. Prodeus, I. Kotvytskyi, «On Reliability of Log-Spectral Distortion Measure in Speech Quality Estimation», *Proceedings of IEEE 5th International Conference Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, P. 17-19. October 2017, Kyiv, Ukraine.
- [75] І. В. Котвицький, А. В. Оса, «Об'єктивне та суб'єктивне оцінювання якості музичних сигналів обмежених смугою частот». *Електронна та акустична інженерія*, 2018, vol. 1, №. 1, С. 34-37.
- [76] J. Blauert, «Group delay distortions in electroacoustical systems»/ Blauert J. // *J. Acoust. Soc. Am.* – Vol.63, No.5. – 1978. – P. 1478-1483.
- [77] A. Prodeus, V. Didkovskyi, M. Didkovska, I. Kotvytskyi, «On Peculiarities of Evaluating the Quality of Speech and Music Signals Subjected to Phase Distortion», *Proceedings of IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO)*, April 18-20, 2017, Kyiv, Ukraine. – P. 455-460.

Додаток 1. Довідки про впровадження результатів роботи

ТОВ «ГЛОБАЛ АКУСТИКС»

03067, м. Київ, вул. Виборзька, 87
Р/р 26009052653625 в ПАТ КБ «ПРИВАТБАНК»
МФО 320649, ЄДРПОУ 40344101

A4 Sound

ЗАТВЕРДЖУЮ
Директор ТОВ Глобал Акустікс
Проказов В.В.
“___” лютого 2020 р.

АКТ ВПРОВАДЖЕННЯ

результатів дисертаційної роботи Котвицького Ігоря Валерійовича
«Оцінювання якості спотворених мовних та музичних сигналів»

на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.09.08 – «Прикладна акустика та звукотехніка»

При виконанні робіт по покращенню акустичних умов в колл-центрі телемагазину «Тор Shop» використані наступні результати дисертаційної роботи:

- 1) Удосконалено метод оцінювання розбірливості мови непрямим методом, через оцінювання якості мовних сигналів:
1. Котвицький І. В., Продеус А. Н. Объективное и субъективное оценивание качества речевых и музыкальных сигналов, подвергнутых фазовым искажениям. - ISSN 1811-4512. Электроника и связь. 2016, № 2, С. 25-31.
2. Котвицький І. В., Продеус А.Н., Дидковский В.С. Об оценивании качества речи с помощью показателя «логарифмические спектральные искажения». Тези доповідей 5-ї міжнародної науково-технічної конференції «Інформаційні проблеми теорії акустичних, радіоелектронних і телекомунікаційних систем» IPST-2016. – Харків: НТУ «ХПІ», 2016, С. 14.
- 2) Запропоновано використання автоматизованої системи оцінювання розбірливості мови артикуляційним методом:
1. І. Котвицький, А. Продеус, А. Вітик, О. Дворник, О. Чайка, М. Ярошенко, Суб'єктивне оцінювання розбірливості мови на тлі шуму та реверберації. "Мікросистеми, електроніка та акустика" том 23, № 2, 2018. ISSN 2523-4447, DOI: 10.20535/2523-4455.2018.23.2.128820, С. 66-73.
2. I. Kotvytskyi, A. Prodeus, V. Didkovskyi, M. Didkovska, "Automated Subjective Evaluation of Speech Intelligibility in Noise and Reverberation," Proceedings of IEEE 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), February 20-24, 2018, Lviv-Slavske, Ukraine, P. 555-558. ISBN: 978-1-5386-2557-6. DOI: 10.1109/TCSET.2018.8336263.

Використання результатів дисертаційної роботи дозволило підвищити надійність прогнозування та вимірювання розбірливості мови на тлі шумової завади та реверберації. Таким чином, результати дисертаційної роботи є науково та практично цінними та можуть бути використані в подальшому при розробці перспективних систем оцінювання якості звукоізоляції різного призначення.

Директор ТОВ Глобал Акустікс

Проказов В.В.



A4 Sound

Тел.: +38 (044) 583-00-94
E-mail: acoustic4sound@gmail.com
www.a4sound.com.ua



ЗАТВЕРДЖУЮ

Перший проректор Національного технічного
університету України "КПІ ім. І. Сікорського"
академік НАН України, професор

Ю. І. Якименко

_____ 2019 р.

ДОВІДКА

про впровадження в учбовий процес кафедри

"Акустика та акустoeлектроніка"

Національного технічного університету України "КПІ ім. І. Сікорського"

результатів дисертаційної роботи Котвицького Ігоря Валерійовича

на тему "Оцінювання якості спотворених мовних та музичних сигналів",

поданої на здобуття наукового ступеню кандидата технічних наук

за спеціальністю 05.09.08 – прикладна акустика та звукотехніка

Наукові положення та результати дисертаційної роботи Котвицького Ігоря Валерійовича на тему "Оцінювання якості спотворених мовних та музичних сигналів" використовуються для викладання навчальних дисциплін студентам, що навчаються за програмами професійного спрямування «Акустичні електронні системи та технології обробки акустичної інформації» спеціальності 171 «Електроніка», а саме: у курсі лекцій з дисципліни «Основи комп'ютерної обробки музики та мови», у розділі «Аналіз мовних, музичних та інших сигналів», у курсі лекцій з дисципліни «Комп'ютерна обробка акустичних сигналів», у розділах «Акустична експертиза комунікаційних каналів», «Автоматизація оцінювання розбірливості мови», а також у курсі лекцій з дисципліни «Захист акустичної інформації», у розділі «Прогнозування та вимірювання розбірливості мови».

Декан факультету електроніки

д.т.н., професор

Жуйков В. Я.

Завідувач кафедри АтаАЕ

д.т.н., професор

Дідковський В. С.

Додаток 2. Таблиця артикуляційних звукосполучень

Таблиця Д1.1. Артикуляційні звукосполучення

няк (няг)	пуль (пуль)	бущ (бущъ)	м'ят (мйад)	свум (свум)	сос (соз)	піц (піц)	чтал (чтал)	дюр (дюр)	хач (хачъ)
мюф (мюв)	зош (зож)	плись (плизь)	риц (риц)	дум (дум)	нись (низъ)	лут (луд)	січ (січъ)	вість (вість)	дис (диз)
фсен (всен)	рай (рай)	бор (бор)	чат (чад)	вис (виз)	гум (гум)	прюх (прюх)	шець (шедь)	тют (тюд)	йоф (йов)
йок (йог)	стял (стял)	тип (тиб)	геф (гев)	бац (бац)	тріт (трід)	шись (шизь)	деп (деб)	сес (сез)	тор (тор)
вох (вох)	жоф (жов)	теф (тев)	трух (трух)	фек (фег)	леф (лев)	вік (віг)	щох (счох)	кась (казъ)	зех (зех)

Додаток 3. Лістинги програм автоматизованої системи суб'єктивного та об'єктивного оцінювання якості та розбірливості мовних та якості музичних сигналів в Matlab

```
%% === Оцінка якості музичного сигналу на виході мікрофонного масиву ===
%
% В даній програмі Microphone_Array_test_istrp.m моделюється ситуація:
% - шум є ізотропним;
% - напрямок на джерело сигналу становить 0 градусів;
% Що робиться в програмі:
% - для SNR=-10дБ, 0дБ, +10дБ оцінюється:
%   1) якість сигналу;
%   2) ступінь придушення завади;
%   3) загальне враження від почутого.
% - при цьому для кожного значення SNR прослуховується 7 сигналів:
%   1) еталонний (незашумлений) музичний сигнал;
%   2) сигнал з виходу одиночного мікрофона;
%   3) сигнал з виходу DS-антени;
%   4) сигнал з виходу LDCM_1-антени;
%   5) сигнал з виходу LDCM_2-антени;
%   6) сигнал з виходу LDCM_3-антени;
%   7) сигнал з виходу LDCM_4-антени;

clc
clear

b=input('Введіть латиницею ваше прізвище (без лапок): ', 's');
disp(' ');
disp('Проводиться група дослідів для ізотропного шуму');
fs = 44100;

for SNR = 10:-10:-10
    tempsig = zeros(1,6); % простір для оцінок якості сигналу
    tempnoise = zeros(1,6); % простір для оцінок рівня завади
    tempwhole = zeros(1,6); % простір для оцінок загального враження
    disp(' ');
    disp(['ВІДНОШЕННЯ СИГНАЛ-ШУМ SNR = ' num2str(SNR)]);
    disp(' ');
    disp('Ви слухаєте ЕТАЛОННИЙ музичний сигнал, його якість 5 балів');
    pause(0.5)
    snd=audioread('cleanmusic.wav');
    sound(snd, fs);

    pause(10)
    g = 5;
    tempsig(1)=g;
    tempnoise(1)=0;
    tempwhole(1)=g;

    disp(' ');
    disp('Прослухайте сигнал 2');
    pause(0.5)
    snd=audioread([num2str(SNR) 'дБ/Array_istrp.wav']);
    sound(snd, fs);
    pause(10)
    g=input('Оцініть якість сигналу (від 1 до 5, можна дробове): ');
    gn=input('Оцініть рівень завади: ');
```

```

gw=input('Оцініть загальне враження від почутого: ');
tempSIG(2)=g;
tempnoise(2)=gn;
tempwhole(2)=gw;

disp(' ');
disp('Прослухайте сигнал 3');
pause(0.5)
snd=audioread([num2str(SNR) 'dB/cbfOut_istrp.wav']);
sound(snd, fs);
pause(10)
g=input('Оцініть якість сигналу (від 1 до 5, можна дробове): ');
gn=input('Оцініть рівень завади: ');
gw=input('Оцініть загальне враження від почутого: ');
tempSIG(3)=g;
tempnoise(3)=gn;
tempwhole(3)=gw;

disp(' ');
disp('Прослухайте сигнал 4');
pause(0.5)
snd=audioread([num2str(SNR) 'dB/LCMVout1_istrp.wav']);
sound(snd, fs);
pause(10)
g=input('Оцініть якість сигналу (від 1 до 5, можна дробове): ');
gn=input('Оцініть рівень завади: ');
gw=input('Оцініть загальне враження від почутого: ');
tempSIG(4)=g;
tempnoise(4)=gn;
tempwhole(4)=gw;

disp(' ');
disp('Прослухайте сигнал 5');
pause(0.5)
snd=audioread([num2str(SNR) 'dB/LCMVout2_istrp.wav']);
sound(snd, fs);
pause(10)
g=input('Оцініть якість сигналу (від 1 до 5, можна дробове): ');
gn=input('Оцініть рівень завади: ');
gw=input('Оцініть загальне враження від почутого: ');
tempSIG(5)=g;
tempnoise(5)=gn;
tempwhole(5)=gw;

disp(' ');
disp('Прослухайте сигнал 6');
pause(0.5)
snd=audioread([num2str(SNR) 'dB/LCMVout3_istrp.wav']);
sound(snd, fs);
pause(10)
g=input('Оцініть якість сигналу (від 1 до 5, можна дробове): ');
gn=input('Оцініть рівень завади: ');
gw=input('Оцініть загальне враження від почутого: ');
tempSIG(6)=g;
tempnoise(6)=gn;
tempwhole(6)=gw;

if SNR == 10
    dB10 = tempSIG;
    dB10n = tempnoise;
    dB10w = tempwhole;

```

```

elseif SNR == 0
    dB0 = tempsig;
    dB0n = tempnoise;
    dB0w = tempwhole;
else
    dBm10 = tempsig;
    dBm10n = tempnoise;
    dBm10w = tempwhole;
end
end

%% запис результатів у файл
d=[b, '_istrp'];
save(d, 'dB10', 'dB0', 'dBm10', 'dB10n', 'dB0n', 'dBm10n', 'dB10w', 'dB0w',
'dBm10w');

%% побудова графіків оцінок якості сигналу
figure
plot(dB10, 'k'); grid; ylabel('DMOS'); xlabel('Тип обробки'); title('Оцінка
якості сигналу, ізотропний');
hold on;
plot(dB0, '--b'); grid; ylabel('DMOS');
hold on;
plot(dBm10, '-.r'); grid; ylabel('DMOS');
legend('SNR = 10 dB', 'SNR = 0 dB', 'SNR = -10 dB');

%% побудова графіків оцінок рівня завади
figure
plot(dB10n, 'k'); grid; ylabel('DMOS'); xlabel('Тип обробки'); title('Оцінка
рівня завади, ізотропний');
hold on;
plot(dB0n, '--b'); grid; ylabel('DMOS');
hold on;
plot(dBm10n, '-.r'); grid; ylabel('DMOS');
legend('SNR = 10 dB', 'SNR = 0 dB', 'SNR = -10 dB');

%% побудова графіків оцінок загального враження
figure
plot(dB10w, 'k'); grid; ylabel('DMOS'); xlabel('Тип обробки'); title('Оцінка
загального враження, ізотропний');
hold on;
plot(dB0w, '--b'); grid; ylabel('DMOS');
hold on;
plot(dBm10w, '-.r'); grid; ylabel('DMOS');
legend('SNR = 10 dB', 'SNR = 0 dB', 'SNR = -10 dB');

%% === table_reading_assist_binaural.m === слухання таблиць ===
% Мета даної програми - автоматизація оцінювання розбірливості
% звукосполучень за результатами прослуховування 3-х наборів (таблиць)
%
% -----
% Кіль-   |
% кість*  |                               Сутність дослідів
% дослідів |
% -----
% 30x3    | Спотворення шумом (БІЛИМ) snr0(dB) = -15; -10; -5; 0; +5
%          | та реверберацією: T60_D = 2.25; 4.0; 5.56; 7.1; 8.68; 10.2
%
%% 1. Вибір режиму роботи:

```

```

clc          % очистка командного окна
clear        % очистка рабочего пространства (workspace)

disp('При ТРЕНУВАННІ прослуховується 3-5 звукосполучень із кожної таблиці');
disp('При РОБОТІ таблиці прослуховуються повністю (50 звукосполучень)');
disp(' ');
wt = input('Оберіть режим роботи? 1 - РОБОТА; 0 - ТРЕНУВАННЯ: '); %!!!
if wt == 0
    numtren = input('Тренування: задайте кількість звукосполучень (від 3 до 5): ');
    nums = numtren; % кількість звукосполучень при тренуванні
else
    nums = 50;
end

%% 2. Запит вхідних даних для assistant_bin.m (прізвище, номери таблиць)
lastname = input('Вкажіть ваше прізвище (без лапок): ', 's');
disp('Вкажіть в порядку зростання номери трьох таблиць для слухання :');
NT = zeros(1,3);
NT(1) = input('Вкажіть номер першої таблиці: N1 = '); %!!!
NT(2) = input('Вкажіть номер другої таблиці: N2 = '); %!!!
NT(3) = input('Вкажіть номер третьої таблиці: N3 = '); %!!!
SNRT = [-15 -10 -5 0 5]; % таблиця значень потрібних SNR0
T60D = [2.25 4.0 5.56 7.1 8.68 10.2]; % дистанції гучномовець-мікрофон

% Створення папки користувача для запису результатів:
kat_results = ['Results_' lastname]; % ім'я основної папки для результатів
mkdir(kat_results); % створення основної папки для запису результатів

% Створ. папки для запису рез-тів досліджень
ngr = 4;
mkdir([kat_results '\Group' num2str(ngr)]);
addpath ('Programs', [kat_results '\Group' num2str(ngr)]); % шлях до програм

clc

%% 3. Запуск програми assistant_bin.m та одержання від неї вихідних даних:
truend = 1;
while (truend == 1)

[params] = assistant_bin_2(kat_results, NT);
N = params.N; % номер еталонної таблиці звукосполучень
m = params.m; % номер групи дослідів (в нас це номер 4)
suf = params.suf;

% 4. Завантаження wav-таблиці із потрібним номером
rfr1_tablename = ['rfr_table_' num2str(N) '_1.mat']; % ім'я 1-ї еталон.таблиці
rfr2_tablename = ['rfr_table_' num2str(N) '_2.mat']; % ім'я 2-ї еталон.таблиці
rfr3_tablename = ['rfr_table_' num2str(N) '_3.mat']; % ім'я 3-ї еталон.таблиці
wav_tablename = ['wav_table_' num2str(N) '_mat']; % ім'я wav-таблиці
wavs = load(wav_tablename); % завантаження wav-таблиці
rzlt_tablename = ['rzlt_table_' suf '.mat']; % ім'я таблиці результатів

% 5. Прослуховування звукосполучення (за натиском довільної клавіші)
clc
disp(['ВИ СЛУХАЄТЕ ТАБЛИЦЮ № ' num2str(N)]);
disp('Якщо треба буде виправити попереднє звукосполучення - введіть 1'); % зміна Чайки
disp(' ');

```

```

if m == 4
    disp('Реверберованийий сигнал + шум')
    disp(['Колір = ' num2str(params.ncolor) ' ; SNR0 = '
num2str(SNRT(params.nsnr)) ...
    ' дБ' ; Distance = ' num2str(T60D(params.nt60)) ' м']]);
    disp(' ');
else
    disp(' ');
end
% === початок спільної модернізації від О.Чайки та М.Ярошенко ===
symb=0;
k=1;
kk=[];
k1=randperm(nums);
while k <=nums
    if symb~='1';
        inputSignal = wavsv(['w' num2str(k1(k))]);
        params.m = m;
        if (m == 2) || (m == 4)
            params.snr0 = SNRT(params.nsnr); % таблиця значень
потрібних SNR0
        end
        if k > 1 && k < nums
            disp('Якщо треба виправити попередній набір - введіть 1'); %
змiна Чайки
        end
        [outputSignal] = distortSystem_bin_50ms(inputSignal, params);
        %pause;
        soundsc(outputSignal,params.Fs);
        textx=horzcat(num2str(k), ' (' ,num2str(k1(k)), '/', num2str(nums), ') Введіть
почуте звукосполучення: ');
        symb = input(textx, 's');
        temprzlt(['s' num2str(k1(k))]) = symb;
        k=k+1;
    elseif symb=='1'
        k=k-1;
        inputSignal = wavsv(['w' num2str(k1(k-1))]);
        params.m = m;
        if (m == 2) || (m == 4)
            params.snr0 = SNRT(params.nsnr); % таблиця значень потрібних
SNR0
        end
        [outputSignal] = distortSystem_bin_50ms(inputSignal, params);
        %pause;
        soundsc(outputSignal,params.Fs);
        textx=horzcat(num2str(k-1), ' (' ,num2str(k1(k-1)), '/', num2str(nums), ')
Кореговане звукосполучення: ');
        symb = input(textx, 's');
        temprzlt(['s' num2str(k1(k-1))]) = symb;
    elseif size(symb,2)==0
        inputSignal = wavsv(['w' num2str(k1(k))]);
        params.m = m;
        if (m == 2) || (m == 4)
            params.snr0 = SNRT(params.nsnr); % таблиця значень
потрібних SNR0
        end
        if k > 1 && k < nums
            disp('Якщо треба виправити попередній набір - введіть 1'); %
змiна Чайки
        end
        [outputSignal] = distortSystem_bin_50ms(inputSignal, params);
        %pause;
        soundsc(outputSignal,params.Fs);

```

```

        textx=horzcat(num2str(k), ' (', num2str(k1(k)), '/', num2str(nums), ') Введіть
почуте звукосполучення: ');
        symb = input(textx, 's');
        temprzlt.(['s' num2str(k1(k))]) = symb;
        k=k+1;
    end
    if k==nums+1 % перевірка останнього звукосполучення
        perevirka = input('Правильно ввели останнє звукосполучення? Так - 1, Ні -
0 ');
        if perevirka==0
            k=k-1;
        end
    end
end
end
% == кінець спільної модернізації від О.Чайки та М.Ярошенко ==

%% 6. Підрахунок частки правильно сприйнятих звукосполучень
rfrtbl_1 = load(rfr1_tablename); % завантаження 1-ї еталон.таблиці
rfrtbl_2 = load(rfr2_tablename); % завантаження 2-ї еталон.таблиці
rfrtbl_3 = load(rfr3_tablename); % завантаження 3-ї еталон.таблиці
num_right = 0; % початкове значення кількості правильних результатів
% clc
%disp('НАОЧНО СПІВСТАВТЕ РЕЗУЛЬТАТ ІЗ ЕТАЛОНАМИ');
%disp(' ');
%disp(['Еталон 1 ' ' Еталон 2 ' ' Еталон 3 ' ' Було введено']);
%disp(' ');
for k1 = 1:nums
    fldname = ['s' num2str(k1)];
    ref1 = rfrtbl_1.(fldname);
    ref2 = rfrtbl_2.(fldname);
    ref3 = rfrtbl_3.(fldname);
    rzlt = temprzlt.(fldname);
    % disp(['ref1 ' ' ref2 ' ' ref3 ' ' rzlt]);
    if ((strcmp(rzlt, ref1))|| (strcmp(rzlt, ref2))|| (strcmp(rzlt, ref3)))
        num_right = num_right + 1;
    end
end
corct = num_right/nums;
disp(' ');
disp(['Правильно сприйнято: ' num2str(num_right) ' із ' num2str(nums)]);
disp(['Оцінка розбірливості становить: ' num2str(corct)]);

% запис відносної кількості правильних результатів в таблицю результатів та
% на диск
temprzlt.('corct') = corct;
katname = [kat_results '\Group' num2str(m) '\ ' rzlt_tablename];
save (katname, '-struct', 'temprzlt');

if params.cnt4 == 90
    disp(' ');
    disp('Поздоровляємо із закінченням першої фази роботи!');
    disp('Зархівуйте папку із результатами та передайте її викладачу. ');
    disp(' ');
    disp('Одержані результати піддайте обробці та підготуйте звіт для захисту
роботи. ');
    return
end

% 7. Запит про подовження слухання таблиць
disp(' ');
truend = input('Подовжити дослід? Так - 1, Ні - 0 ');
end

```


Додаток 4. Свідоцтво про авторське право на комп'ютерну програму

УКРАЇНА



СВІДОЦТВО
про реєстрацію авторського права на твір

№ 85027

Комп'ютерна програма "Автоматизована система суб'єктивного оцінювання розбірливості мови"

(вид, назва твору)

Автор(и) Продеус Аркадій Миколайович, Вітик Андрій Віталійович, Дворник Олександр Олександрович, Котвицький Ігор Валерійович, Чайка Олександр Сергійович, Ярошенко Михайло Олексійович

(повне ім'я, псевдонім (за наявності))

Дата реєстрації 30.01.2019

Державний секретар Міністерства економічного розвитку і торгівлі України О. Ю. Перевезенцев

